

SATURNO V 6.11

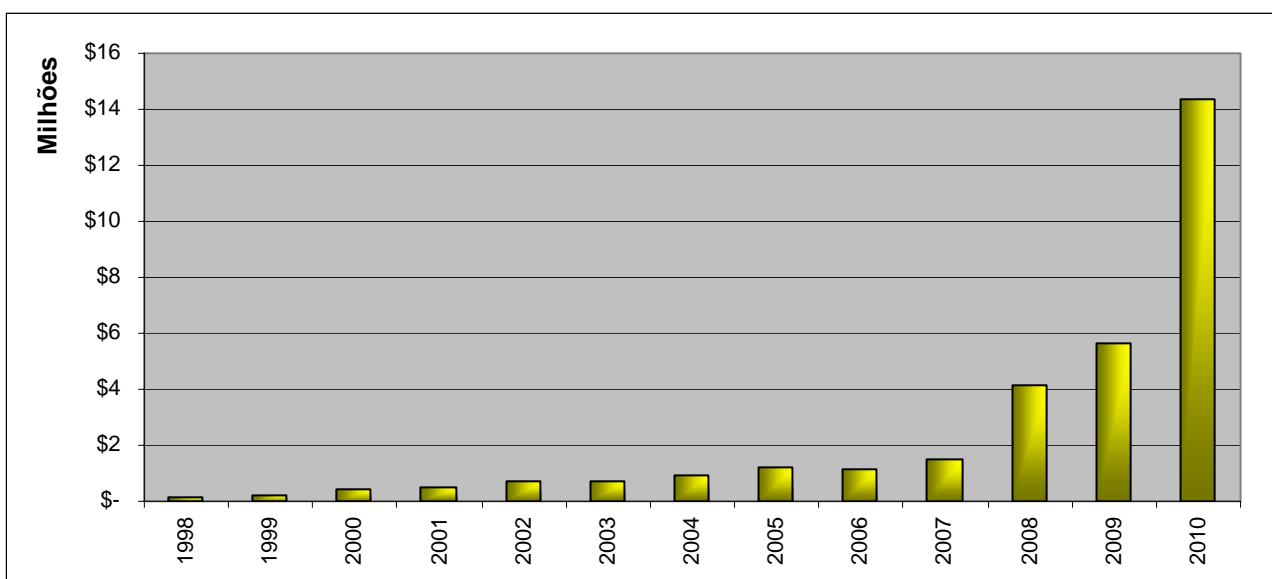
Por Hindenburg Melão Jr.

<http://www.saturnov.com>

A versão 6.11 (ou 3.1415926c83) foi analisada sob diversos aspectos, a fim de verificar a uniformidade do comportamento ao longo do tempo. Primeiramente foram feitas otimizações a fim de selecionar uma configuração para testes mais detalhados. No final, duas configurações se mostraram particularmente interessantes: uma final de dia desde 1973 e outra intradiária desde julho de 1998. Esta segunda se mostrou mais interessante para uma primeira análise.

Um dos resultados interessantes foi que ela, otimizada entre julho de 1998 e julho de 1999, mantendo a mesma configuração, conseguiu produzir mais de 1300% de lucro entre 1999 e 2010, com máximo drawdown de 73%. Otimizada nos últimos 3 meses (7/5/2010 a 7/8/2010) conseguiu produzir lucros desde 2000, com total de quase 600% no período e máximo drawdown de 64%. Não é tão bom um resultado com 1300% e drawdown 72,7% ou 600% e drawdown de 64%, mas estas não são melhores otimizações. São otimizações no limite extremo de curtíssimo período, para validação em período muitíssimo mais longo, para verificar a uniformidade da versão ao longo do tempo, e se pequenos intervalos de tempo reúnem informações suficientes sobre as propriedades do Mercado, para produzir valores para os parâmetros que se mantenham válidos por períodos muito mais longos. Constatamos que isso não é aplicável todas as vezes, como acontece nas configurações para final de dia. Com as configurações intradiárias, isso só se observa em alguns intervalos específicos.

A melhor otimização intradiária produziu 14.300% de lucro entre 1998 e 2010, com máximo drawdown de 42%, e um total de quase 4.000 operações realizadas. Isso representa um lucro médio de mais de 50% ao ano durante 12 anos e transforma um investimento inicial de \$ 100.000,00 em \$ 14.400.000,00, num período que o DJI teria transformado os mesmos \$ 100.000,00 em meros \$ 117.000,00, além disso, o máximo drawdown no DJI no período foi de -55%, enquanto o máximo drawdown no Saturno V 6.11 foi -42%, e a máxima perda anual do Saturno V no período foi -3%, enquanto a máxima perda anual do DJI foi perto de -50%. Essa rápida recuperação se deve à possibilidade de o Saturno V inverter rapidamente a posição quando a tendência se inverte, enquanto o DJI não tem como reverter essa situação mesmo com assistência externa de quase 1 trilhão de dólares para tentar segurar artificialmente os preços das ações.



As tabelas a seguir resumem alguns resultados relevantes do Saturno V 6.11:

Início	Fim	Balanço	Trades	T/ano	Lucro acumulado	Lucro anual	Média anual
10/07/1998	31/12/1998	\$ 156,624.27	145	304	56.62%	56.62%	156.47%
01/01/1999	31/12/1999	\$ 196,030.12	496	351	96.03%	96.03%	57.79%
01/01/2000	31/12/2000	\$ 407,979.40	827	331	307.98%	108.12%	76.38%
01/01/2001	31/12/2001	\$ 525,877.93	1,137	310	425.88%	28.90%	61.19%
01/01/2002	31/12/2002	\$ 709,643.12	1,416	279	609.64%	34.94%	54.92%
01/01/2003	31/12/2003	\$ 740,580.23	1,735	319	640.58%	4.36%	44.15%
01/01/2004	31/12/2004	\$ 924,260.03	2,054	319	824.26%	24.80%	40.96%
01/01/2005	31/12/2005	\$ 1,204,613.65	2,353	299	1,104.61%	30.33%	39.49%
01/01/2006	31/12/2006	\$ 1,165,458.43	2,649	296	1,065.46%	-3.25%	33.60%
01/01/2007	31/12/2007	\$ 1,487,389.19	2,985	336	1,387.39%	27.62%	32.96%
01/01/2008	31/12/2008	\$ 4,173,351.63	3,332	347	4,073.35%	180.58%	42.78%
01/01/2009	31/12/2009	\$ 5,656,475.44	3,694	362	5,556.48%	35.54%	42.13%
01/01/2010	07/08/2010	\$ 14,333,876.13	3,897	340	14,233.88%	153.41%	50.85%

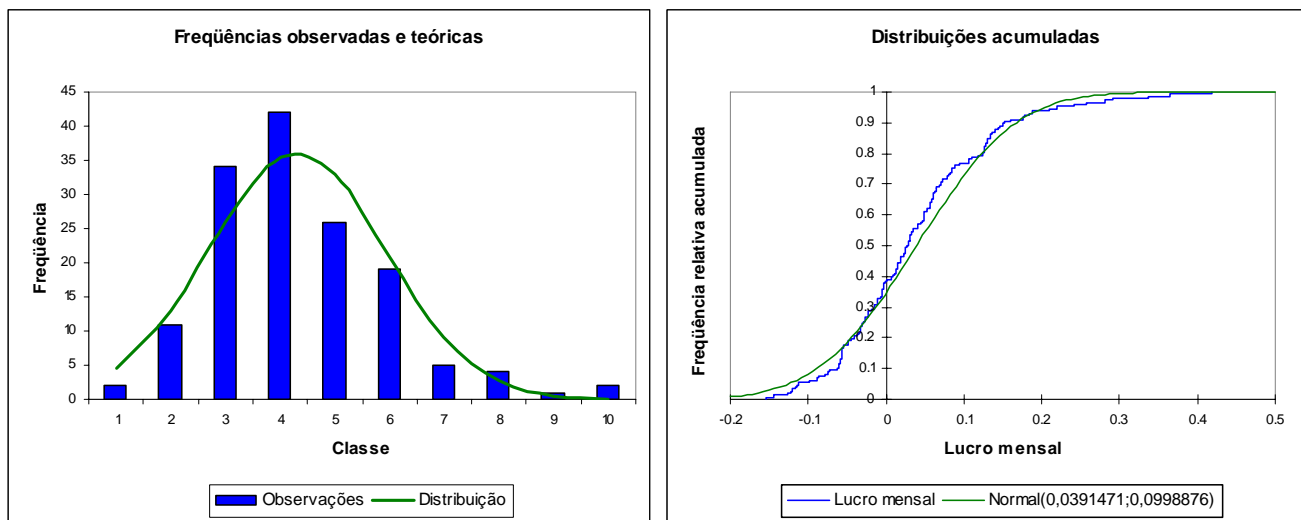
Mínimo lucro anual	-3.25%
Máximo lucro anual	180.58%
Média aritmética dos lucros anuais	59.85%
Média anual	50.85%
Tempo transcorrido (anos)	12.08
Média de trades por anos	322.69
Fator de lucro	1.30
Fator de lucro anualizado	240.36
Porcentagem de trades lucrativos	37.68%
Operações com lucro	1457
Operações com prejuízo	2410

Outra propriedade que nos interessa é a correlação ordinal de Spearman entre o tempo transcorrido, em meses, e o lucro auferido até aquele momento: 0,981. Os períodos foram estratificados de 1 em 1 mês, e essa alta correlação mostra que há muita constância no crescimento. Porém essa informação não diz muito sobre o tamanho das quedas em comparação ao tamanho dos ganhos em diferentes períodos, nem sobre a uniformidade no crescimento. A correlação produto-momento de Pearson entre o tempo transcorrido e o logaritmo do lucro foi 0,934, mostrando que o tamanho dos ganhos em comparação ao tamanho das perdas se manteve muito semelhante ao longo do tempo, bem como mostrando que o ritmo de crescimento é suave e bastante uniforme na maior parte do tempo. A correlação de Pearson ponderada em função do inverso do tempo retroativo foi 0,946, indicando que com o passar do tempo a correlação entre tempo transcorrido e lucro acumulado está se tornando mais forte, por isso a correlação ponderada temporalmente é um pouco maior que a correlação sem ponderação. Na verdade, indica que recentemente está se tornando mais forte. Quando a ponderação é em função do próprio tempo, em vez de o inverso do tempo retroativo, a correlação cai para 0,920. Isso pode ser interpretado como uma redução aproximadamente linear da correlação ao longo do tempo, mas um aumento exponencial da correlação ao longo do tempo, sendo que a redução linear se impõe no início do período e vai progressivamente sendo superada pelo aumento exponencial. Mas o intervalo é demasiado curto para que se possa inferir com segurança algo nesse sentido.

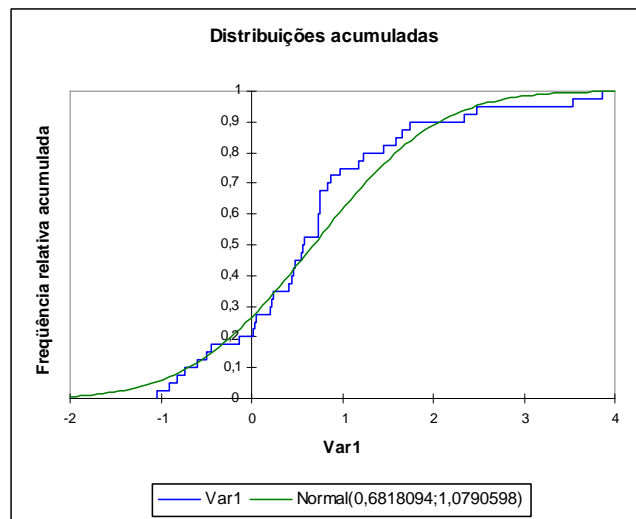
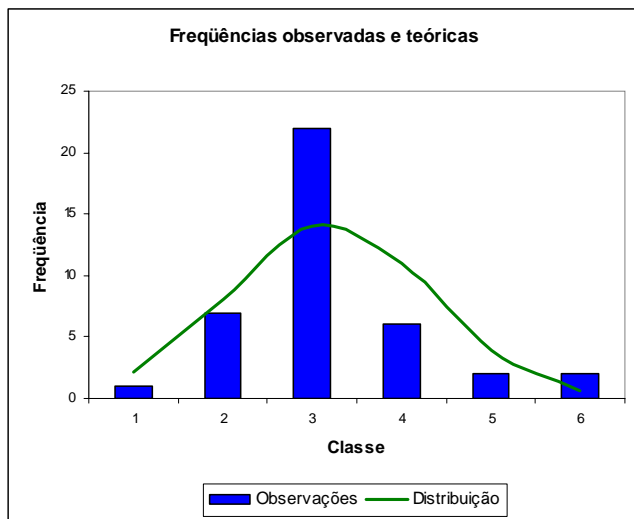
A grande quantidade de operações, de quase 4.000 trades, possibilita realizar vários estudos que não seriam possíveis com amostras menores, por isso aproveitamos a oportunidade para investigar algumas propriedades da estratégia desta configuração e da série histórica, entre outras. Para tanto, foram feitas duas Análises de Conglomerados Univariadas, uma com 2 clusters e outra com 5 clusters. Antes, porém, realizamos testes qualidade de ajuste de Kolmogorov-Smirnov e Chi-quadrado para conferir se deveria ser rejeitada a hipótese de que os lucros mensais apresentam distribuição normal. Com base no teste de Chi-quadrado, a

maior sensibilidade na cauda resultou na rejeição da hipótese nula, enquanto Kolmogorov-Smirnov, mais sensível à forma, a hipótese não foi rejeitada.

Mas um teste de qualidade de ajuste não diz nada sobre quão semelhante é uma distribuição empírica em comparação a uma teórica. Diz apenas qual a probabilidade, com base na dispersão das diferenças medidas, de que uma distribuição empírica não possa ser representada por determinada distribuição teórica. Isso significa que quanto maior a amostra de dados, menor será a dispersão nos dados e maior a probabilidade de se encontrar alguma diferença sensível entre as distribuições empírica e teórica. Quando a amostra é muito pequena, mesmo que os histogramas se encaixem muito grosseiramente nas curvas teóricas, os testes de qualidade de ajuste acabam levando a não rejeitar a hipótese nula, ou seja, mesmo havendo grandes diferenças entre a forma e as propriedades gerais da distribuição teórica e a empírica, se a amostra for pequena, portanto com grandes incertezas nas medidas, os testes acabam não descartando hipóteses ruins, ao passo que para amostras grandes, mesmo quando os dados empíricos se encaixam muito melhor nos modelos teóricos, pode ocorrer de excelentes ajustes serem rejeitados, porque como a precisão aumenta com o tamanho da amostra, pequenas diferenças acabam sendo suficientes para indicar que a distribuição teórica não é aderente o bastante para representar a distribuição empírica dentro dos estreitos limites de incerteza que a grande amostra proporciona. Os dois gráficos abaixo mostram um ajuste de uma normal aos dados empíricos dos lucros mensais:

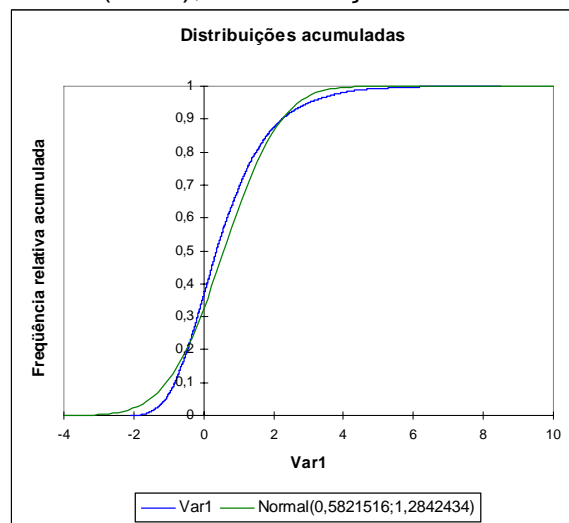
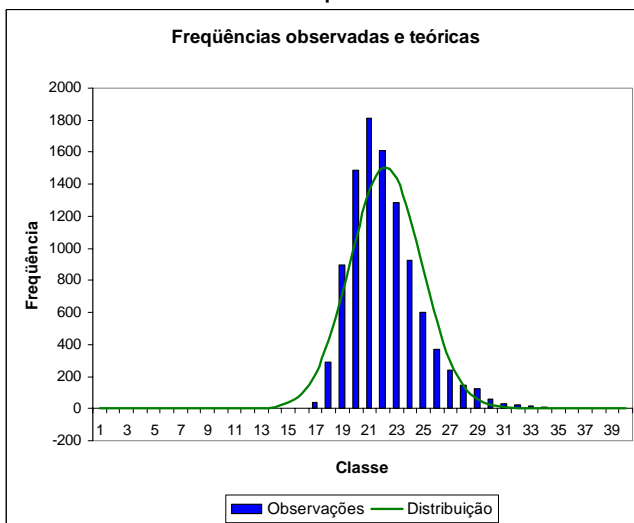


Principalmente no segundo gráfico, que mostra a probabilidade cumulativa, vemos que o ajuste é bom. E como a amostra tem 146 elementos, é suficientemente grande para que haja hipersensibilidade nos testes de qualidade de ajuste, levando a rejeitar incorretamente uma distribuição satisfatória. Apenas como exemplo, para corroborar esse fato, vejamos duas distribuições geradas pelo Método de Monte Carlo, analisadas com os mesmos testes de qualidade de ajuste. A primeira distribuição é uma Gumbel, bem diferente de uma gaussiana, porém com apenas 40 elementos.



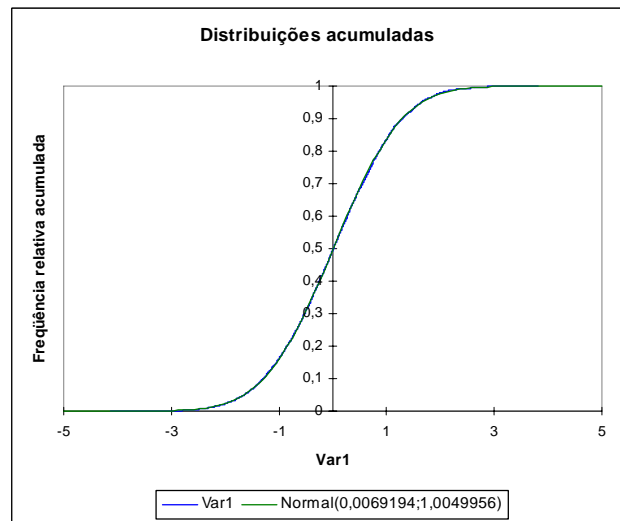
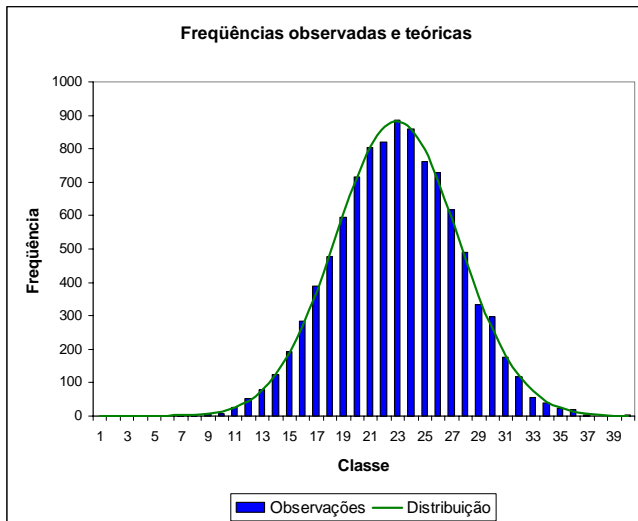
A linha verde da normal à esquerda está meio torta porque o recurso “suavizar” do Excel não leva em conta qual função ele está suavizando

A disparidade entre a normal e os dados pseudo-empíricos é visível e a qualidade de ajuste é bem pior do que no caso anterior. Enfim, podemos observar que a distribuição teórica não representa bem os dados empíricos, mas como a amostra é muito pequena, e a incerteza muito grande, tanto o teste de Shapiro-Wilk quanto o de Kolmogorov-Smirnov levam à não rejeição de que estes dados empíricos tenham distribuição normal. Se essa amostra de 40 elementos fosse ampliada para 10.000, usando mesma semente no Mersenne Twister e mantendo os mesmos parâmetros da Gumbel acima (0 e 1), a distribuição ficaria assim:



Claro que nesse caso tanto Kolmogorov-Smirnov quanto Chi-quadrado rejeitam a hipótese de que os dados representados pelas barras azuis tenham distribuição normal, porque a precisão aumentou muito e a diferença entre a curva teórica e as barras pseudo-empíricas se tornou muito mais nítida como algo “real”, em vez de uma flutuação causada por ruído. A assimetria na Gumbel, que vai se delineando mais claramente à medida que a amostra cresce, torna impossível ajustar uma amostra tão grande a uma normal, mas numa amostra com poucos elementos a assimetria se mistura aos ruídos e fica difícil determinar o que é assimetria e o que é ruído, aumentando a incerteza e alargando a margem para não rejeição de ajustes ruins.

Agora vejamos uma distribuição “de fato” gaussiana (o Systat 12 foi configurado para gerar 10 distribuições normais com média 0 e desvio-padrão 1, pelo Método de Monte Carlo). Se houver muitos elementos (10.000 nesse caso), então uma distribuição teórica com ajuste excepcionalmente bom pode ser rejeitada para modelar a distribuição pseudo-empírica que tenha sido produzida seguindo uma distribuição normal, como podemos ver abaixo:

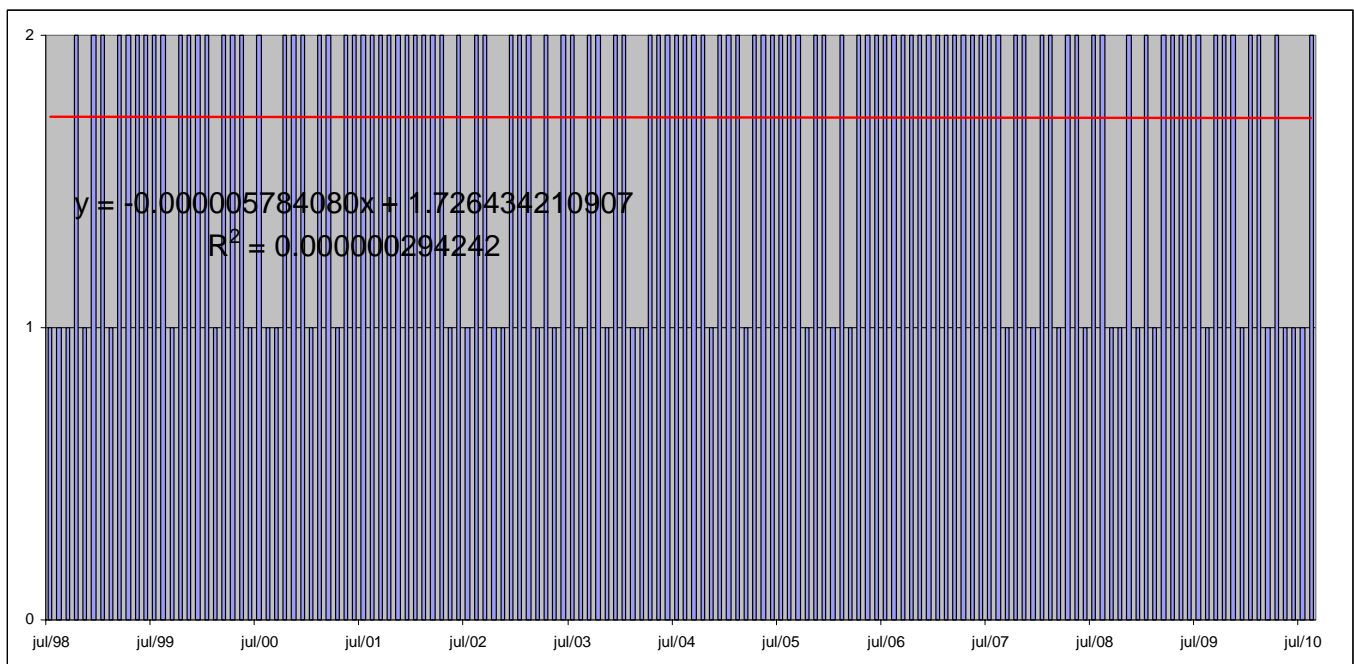


A aderência da curva aos dados pseudo-empíricos é “linda”, quase perfeita na distribuição de densidade de probabilidades, à esquerda, e mal dá pra notar que existem duas curvas (uma azul e uma verde) na cumulativa à direita. Mesmo assim, devido a pequenas disparidades principalmente nas caudas, que nem são perceptíveis visualmente, somadas à pequena distorção próxima ao topo, o p-valor fica abaixo de 0,05 e a hipótese de distribuição normal é rejeitada. Apenas 1 entre as 10 distribuições normais com 10.000 elementos geradas pelo Método de Monte Carlo foi rejeitada, mesmo assim é um absurdo que uma distribuição com tanta aderência como esta seja rejeitada, enquanto outra tão pouco aderente, como a de Gambel com 40 elementos, não tenha sido rejeitada. Isso acontece simplesmente porque quando a amostra é muito grande, a precisão também é grande, e quanto maior a precisão, mais o teste de qualidade de ajuste é sensível a pequenas diferenças.

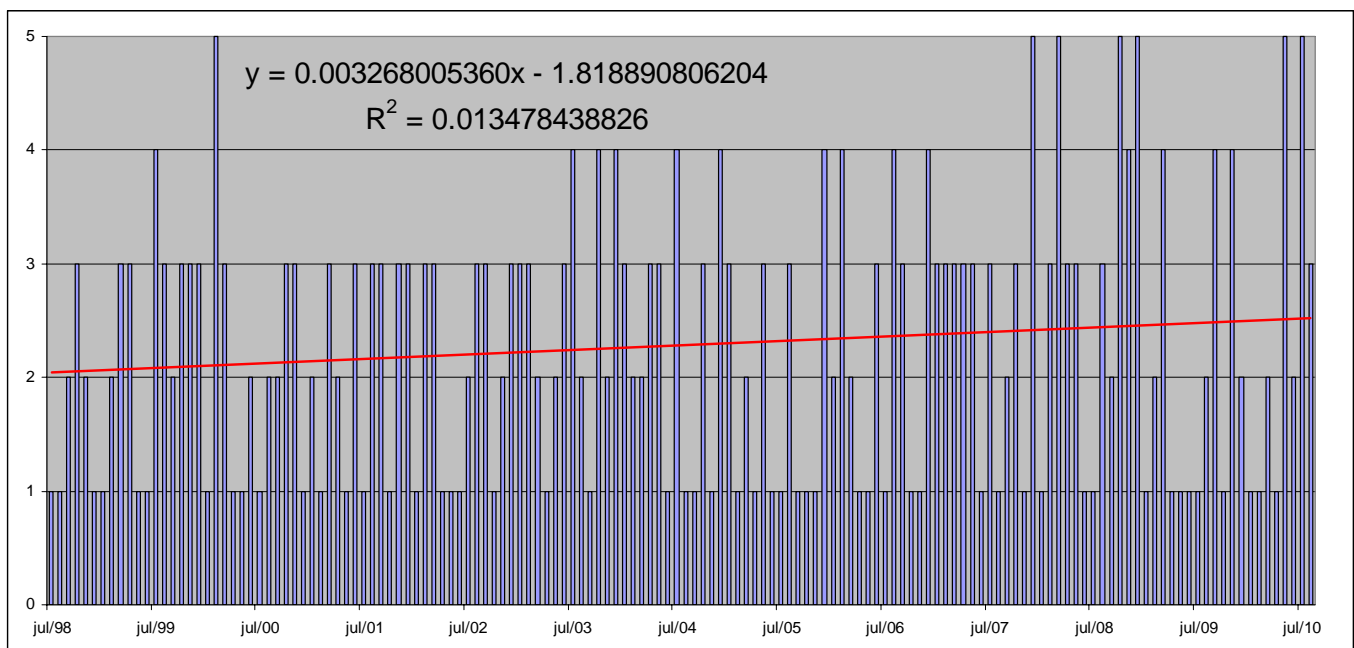
Para nosso caso, é mais útil que a forma da distribuição empírica seja semelhante à teórica, não importando a probabilidade de que as diferenças observadas sejam estatisticamente significativas, logo não é tão relevante que um teste de Chi-quadrado rejeite a hipótese nula. Por esses motivos, consideramos apropriado aceitar a hipótese de que os lucros mensais se distribuem normalmente.

Por isso também é que, embora seja praxe realizar testes de qualidade de ajuste, antes de outros testes que necessitem da validação da hipótese de normalidade, muitas vezes prefiro nem realizar tais testes, já que os resultados podem acabar sendo irrelevantes para decisões subsequentes.

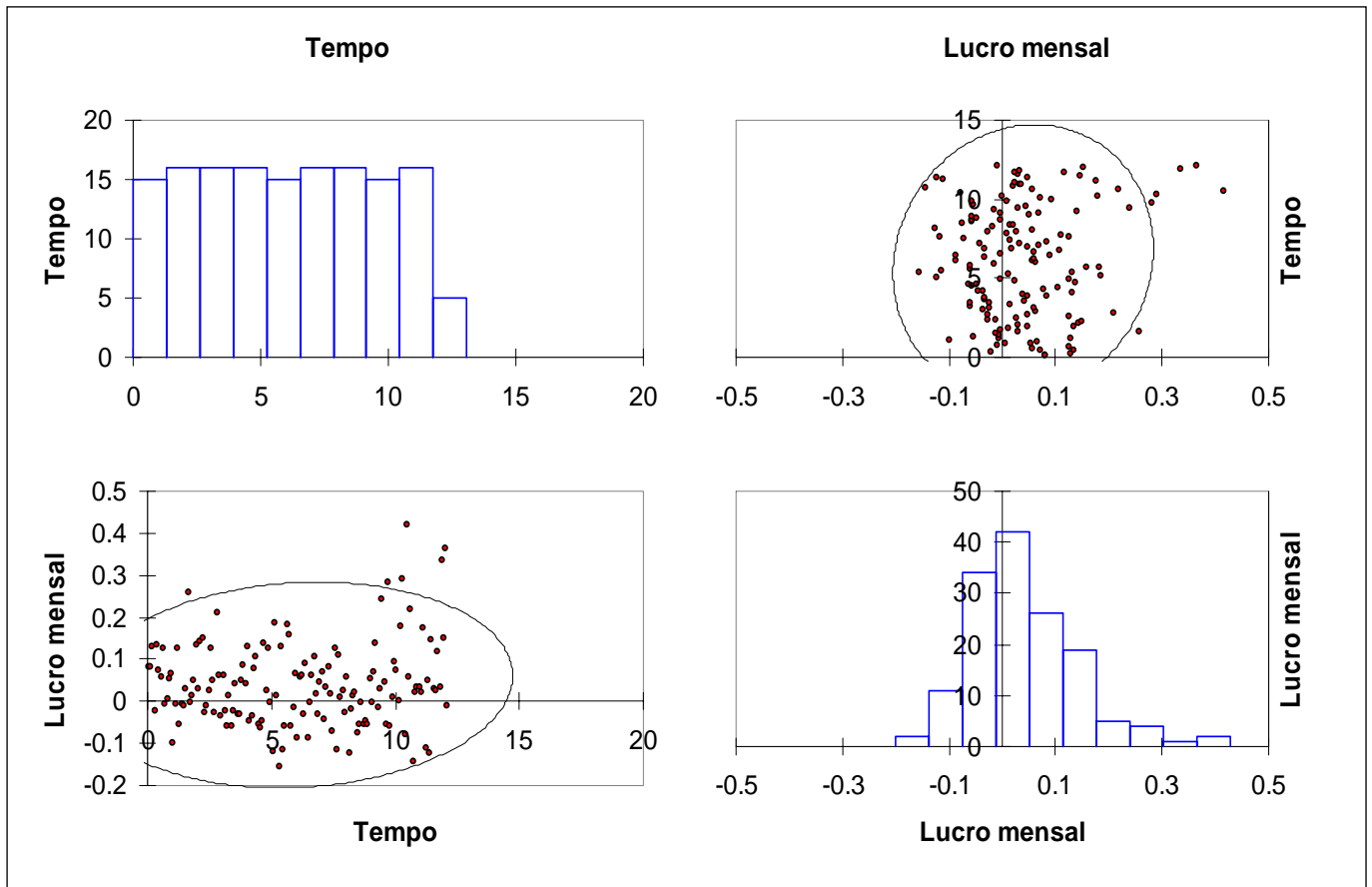
Enfim, no caso da Análise de Clusters com 2 classes, constatamos que os elementos nos dois clusters se distribuem muito uniformemente ao longo do tempo, de modo que uma regressão linear produz uma reta (vermelha) praticamente paralela ao eixo das ordenadas, conforme podemos ver no gráfico abaixo, em que a altura das barras, 1 ou 2, indica o cluster em que cada mês ficou classificado, e o eixo x mostra o tempo em meses:



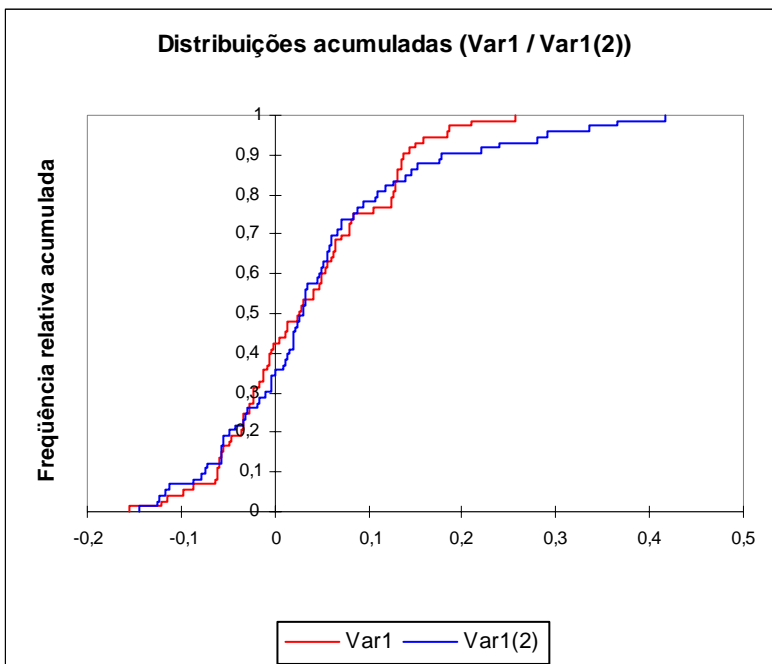
A Análise de Clusters com 5 classes também mostrou uma distribuição muito uniforme, e com ligeira inclinação da reta, mostrando predominância dos clusters classificados na classe 5 entre os meses mais recentes. Essa característica poderia ser incorretamente interpretada como uma tendência de que os lucros estejam se tornando cada vez maiores, com o passar do tempo. Mas isso não pode ser inferido com base apenas nisso, pois também os clusters de classe 4 estão predominando entre os dados mais recentes, e na classe 4 estão os meses com centróide -11%, portanto os que resultaram em maior prejuízo. O gráfico abaixo resume isso:



O correto para determinar se os lucros estão aumentando ou não ao longo do tempo é uma correlação produto-momento de Pearson entre o tempo transcorrido e os lucros mensais. O resultado é 0,090, bastante baixa e não é estatisticamente significativa a um nível 0,05. Além disso, parte dessa correlação é produzida por outliers em 2008 e 2010.



Também realizamos outros testes para verificar a homogeneidade, homoscedasticidade, consistência e outras propriedades da distribuição dos lucros mensais em função do tempo. Alguns dos resultados estão na planilha Excel que pode ser baixada no link ao final do artigo. Quase todos os testes realizados sugerem não haver diferença significativa nas performances em função do tempo, quando se toma o tempo em intervalos grandes, mas podem ocorrer diferenças punctuais, como houve em 2006 e 2007, inclusive 2006 foi o único ano negativo. O teste de Bartlett sugere que a variância nos primeiros 6 anos foi significativamente diferente da observada nos últimos 6 anos, o que não constitui necessariamente um problema, já que grande parte desse efeito está associada ao fato de que os ganhos têm sido maiores nos últimos anos. Dividindo em intervalos de 2 em 2 anos, o resultado é mais marcante,

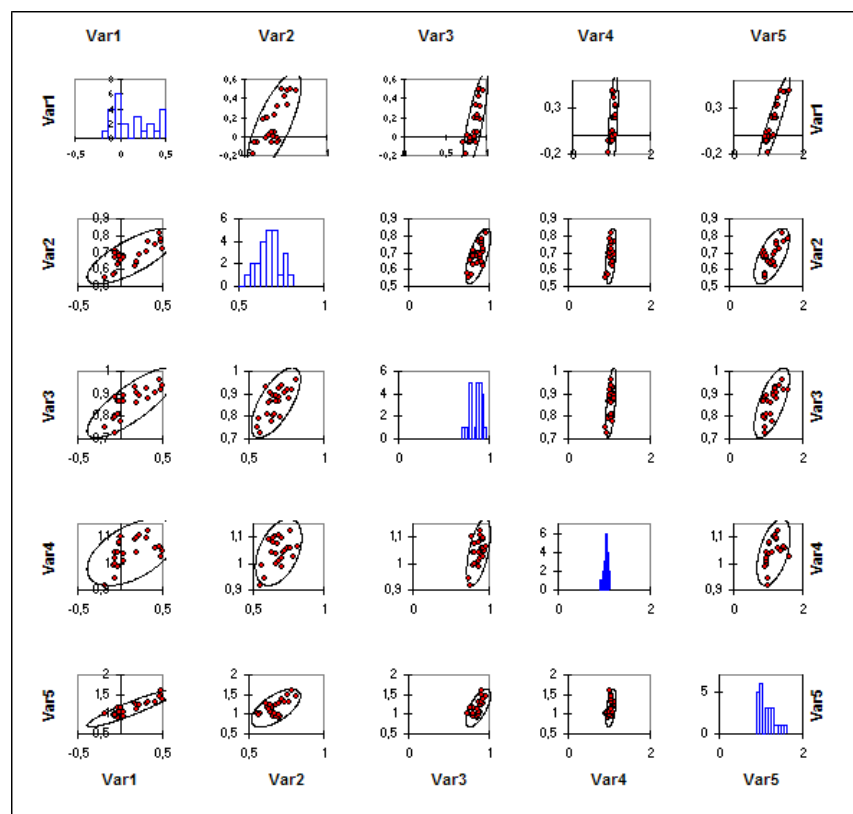


corroborando que a variância na carteira de fato não é a mesma ao longo do tempo. O teste de Kolmogorov-Smirnov comparando os primeiros 73 meses com os últimos 73 meses também mostrou uma distribuição muito similar, em que a diferença principal se deve aos outliers recentes. Isso também ajuda a explicar porque a variância não é uniforme ao longo do tempo.

O teste de Mann-Whitney entre os primeiros e os últimos 73 meses corrobora este resultado, com p-valor muito semelhante ao obtido por meio de Kolmogorov-Smirnov: 0,839 contra 0,894. Com Kruskal-Wallis para 5

períodos de 24 meses cada, o p-valor foi 0,712.

A correlação produto-momento de Person entre logaritmo do lucro acumulado, com agrupamentos de 2 em 2 anos, também se mostrou significativa em todas as 10 comparações, variando entre 0,431 e 0,926 (abaixo). A correlação corrigida (semelhante a correção para Split-Half) passa ser de 0,849 a 0,987. Portanto todas as correlações são extremamente fortes.



A multicolinearidade dos lucros acumulados, considerados de 2 em 2 anos, também apresentou resultados similares. O α de Cronbach para lucros acumulados agrupados em 12 períodos de 12 meses cada foi 0,834, muito semelhante ao valor obtido com split-half com correção de Spearman-Brown: 0,803.

A conclusão a que somos levados é de que há abundância de evidências de uniformidade e estabilidade nas performances ao longo do tempo. Infelizmente não há como fazer testes igualmente detalhados com outras configurações porque a quantidade de operações realizadas fica entre algumas dezenas e poucas centenas. Nesse caso, com 3.900 operações é possível dividir estas operações em grupos e comparar as propriedades dos grupos entre si.

Não realizamos os mesmos testes para verificar se a quantidade de operações em função do tempo apresenta variação porque a correlação entre a quantidade de operações acumulada e o tempo transcorrido é $r=0,99971$.

Download da planilha Excel:

http://www.saturnov.com/contas/imagens/Saturno_611.rar