

Avaliando a diferença entre os anos de hanseníase no Brasil

Mikael Marin Coletto

2024-12-12

1 Introdução

A hanseníase é uma doença infecciosa causada pelo bacilo *Mycobacterium leprae*. Embora tenha sido controlada em muitas partes do mundo, ainda persiste em algumas regiões, especialmente em países em desenvolvimento como o Brasil, Índia e Indonésia. A Organização Mundial da Saúde (OMS) estabeleceu metas para eliminar a hanseníase como problema de saúde pública, definindo uma taxa de incidência anual inferior a 1 caso por 10.000 habitantes. No contexto brasileiro, onde a doença ainda apresenta uma considerável prevalência, os governos federais tem tentado implementar estratégias para aumentar a vigilância e consequentemente facilitar o tratamento, já que é uma doença de tratamento relativamente simples se diagnosticada em estágios iniciais.

Neste relatório, utilizaremos testes estatísticos não paramétricos para avaliar as diferenças nas prevalências de hanseníase no Brasil ao longo dos anos, comparando diferentes estados. Além disso, analisaremos a diferença no número de casos entre dois estados específicos para determinar a eficácia das políticas públicas de prevenção da doença.

2 Métodos

Para esta análise, coletamos dados oficiais de casos de notificação de hanseníase nos estados brasileiros ao longo de vários anos. Os dados foram obtidos do Ministério da Saúde do Brasil, usando o tabnet e baixando a base do Sistema de informação de Agravos de Notificação (SINAN) e abrangem um período de 23 anos (2001-2023).

Inicialmente faremos uma análise descritiva da base de dados.

Primeiramente, realizamos um teste não paramétrico para duas amostras independentes para comparar a diferença na incidência de hanseníase entre estados ao longo dos anos.

Especificamente, utilizamos o teste de Mann-Whitney U, que é adequado para comparar medias de duas amostras independentes quando as suposições dos testes paramétricos não são satisfeitas.

Além disso, selecionamos dois estados com diferentes abordagens em políticas públicas de tratamento da hanseníase e compararamos o número de casos nestes estados ao longo do mesmo período de tempo. Novamente, aplicamos o teste de Mann-Whitney U para determinar se há uma diferença significativa nos números de casos entre esses dois estados.

3 Metodologia

4 Resultados

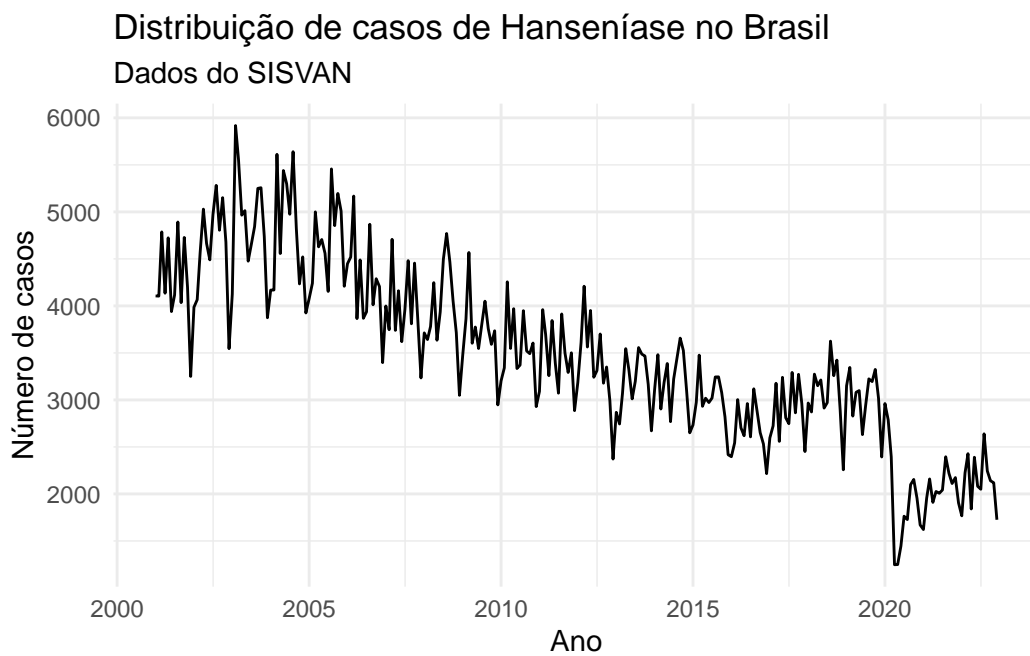
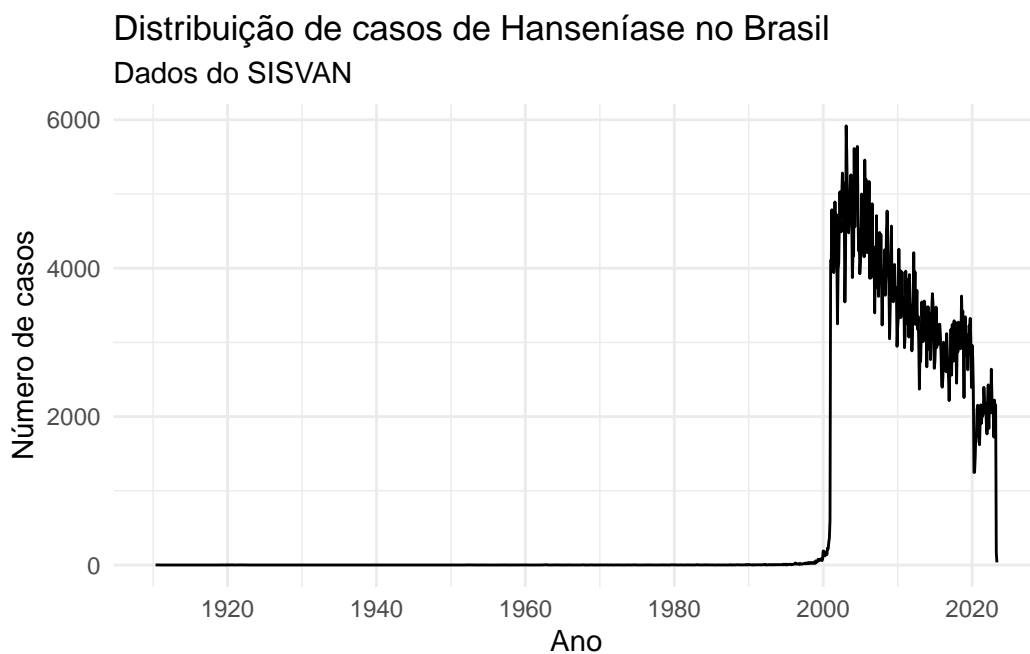
4.1 Avaliando banco de dados

A primeira etapa desenvolvida foi uma análise exploratória dos dados de hanseníase obtidos através do sinan, para isso, carregamos os dados e observamos a completude dos dados no quesito data de diagnóstico e o ID do agravo (CID).

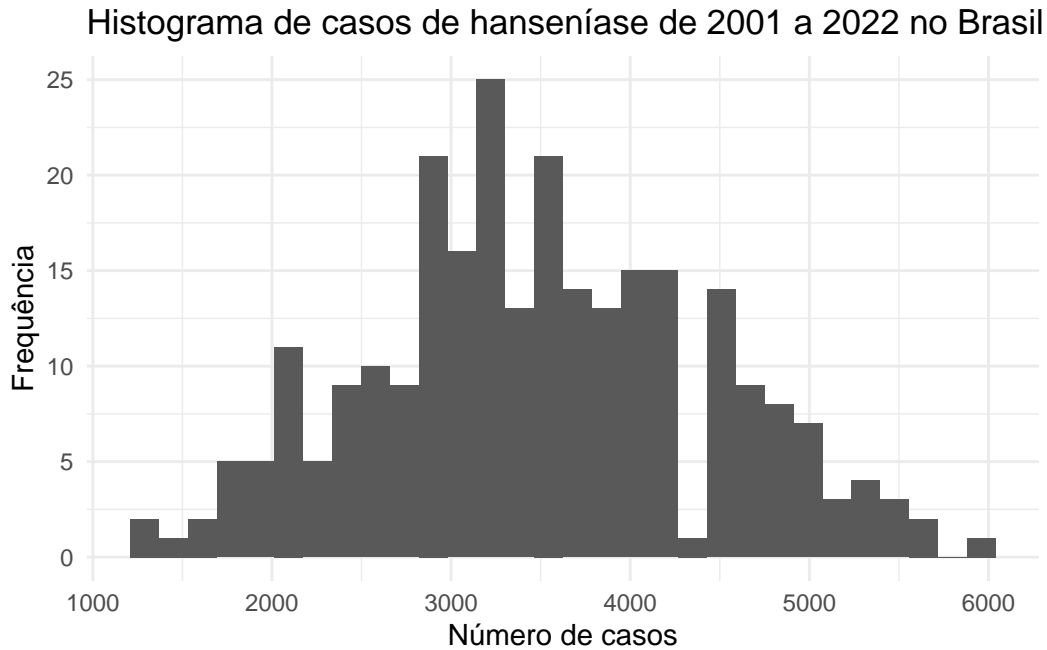
Pelo código desenvolvido, foi observado que o único CID disponível nesta base de dados é o de Hanseníase A309. Além disso, não temos nenhum caso sem CID ou sem data de diagnóstico preenchida (a contagem retornou 0). Porém, temos casos de diagnóstico em datas muito mais antigas do que a análise propôs, portanto os casos anteriores à 2001 serão removidos da base de dados logo adiante, o que pode ser um erro de preenchimento ou um caso de diagnóstico tardio.

4.2 Análise exploratória

4.2.1 Analisando o número de casos por ano/mês no Brasil



4.2.2 Analisando normalidade dos dados



Shapiro-Wilk normality test

```
data:  sinan_hans_2001_2023_anomes_brasil$n
W = 0.99287, p-value = 0.2385
```

```
Warning in ks.test.default(sinan_hans_2001_2023_anomes_brasil$n, "pnorm", :
ties should not be present for the Kolmogorov-Smirnov test
```

Asymptotic one-sample Kolmogorov-Smirnov test

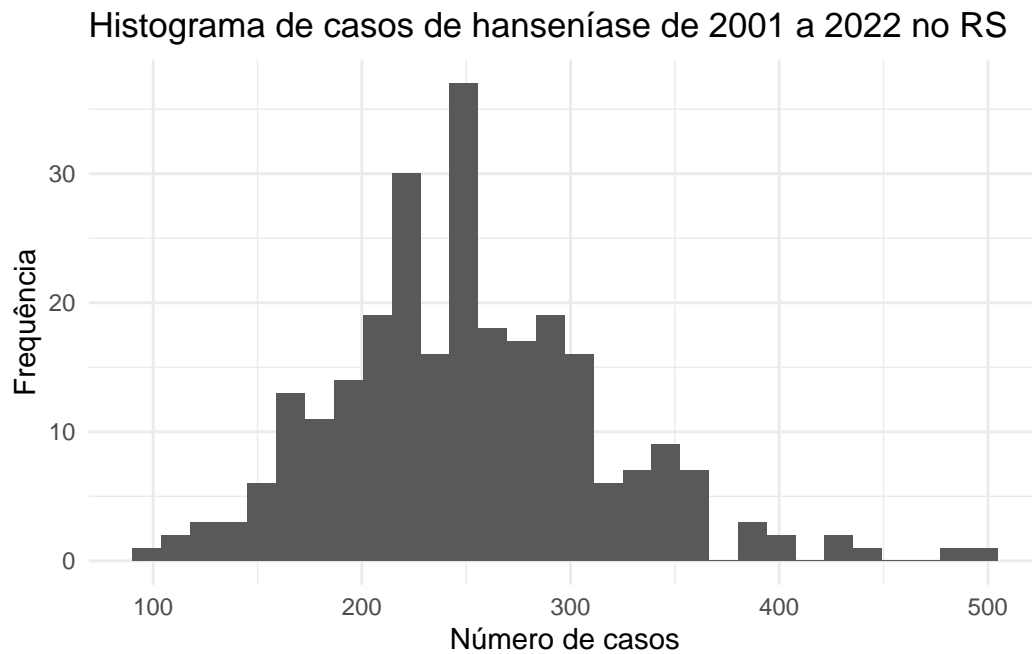
```
data:  sinan_hans_2001_2023_anomes_brasil$n
D = 0.041858, p-value = 0.744
alternative hypothesis: two-sided
```

Os dados de hanseníase para o Brasil de 2001 a 2022 parecem seguir uma distribuição normal, como indicado pelo teste de Shapiro-Wilk (p-valor > 0.05), o teste de Kolmogorov-Smirnov (p-valor > 0.05) e observando o histograma.

4.2.3 Analisando dados por UF

Agora iremos observar os dados de algumas UFs, para isso, iremos agrupar os dados por UF e ano/mês e observar a distribuição de casos de hanseníase para Bahia e Rio Grande do Sul.

``summarise()`` has grouped output by 'UF'. You can override using the `` .groups `` argument.



Shapiro-Wilk normality test

```
data:  dplyr::pull(sinan_hans_uf_2001_2022_ba, n)
W = 0.97844, p-value = 0.0004961
```

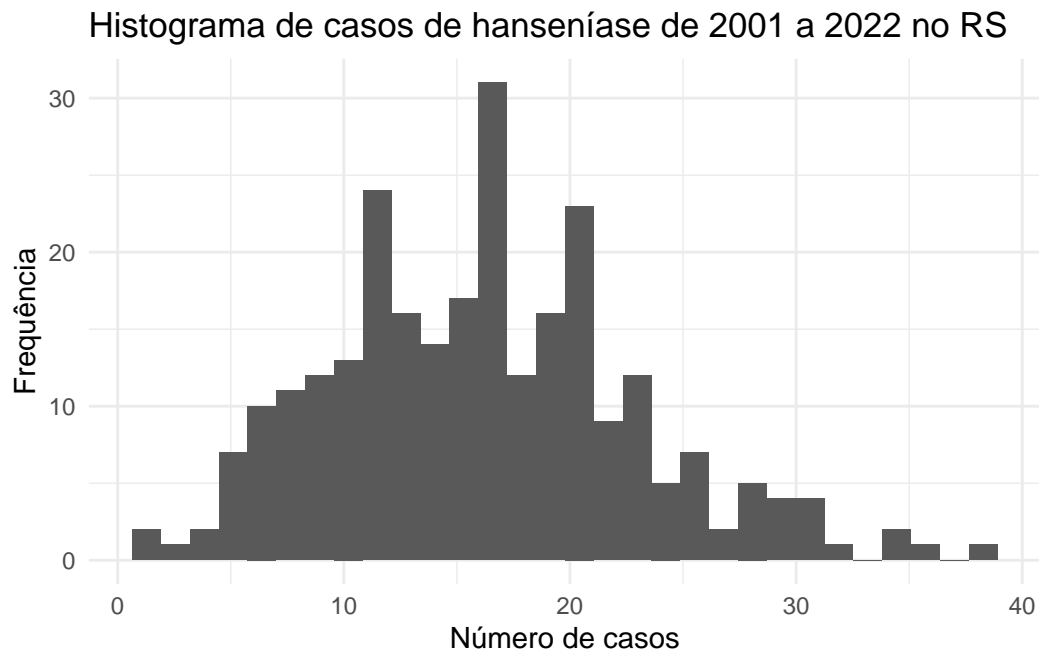
```
Warning in ks.test.default(sinan_hans_uf_2001_2022_ba$n, "pnorm",
mean(sinan_hans_uf_2001_2022_ba$n), : ties should not be present for the
Kolmogorov-Smirnov test
```

Asymptotic one-sample Kolmogorov-Smirnov test

```
data:  sinan_hans_uf_2001_2022_ba$n
D = 0.068696, p-value = 0.1654
alternative hypothesis: two-sided
```

Anderson-Darling normality test

```
data:  sinan_hans_uf_2001_2022_ba$n
A = 1.1776, p-value = 0.004416
```



Shapiro-Wilk normality test

```
data:  dplyr::pull(sinan_hans_uf_2001_2022_rs, n)
W = 0.98537, p-value = 0.008451
```

```
Warning in ks.test.default(sinan_hans_uf_2001_2022_rs$n, "pnorm",
mean(sinan_hans_uf_2001_2022_rs$n), : ties should not be present for the
Kolmogorov-Smirnov test
```

Asymptotic one-sample Kolmogorov-Smirnov test

```
data:  sinan_hans_uf_2001_2022_rs$n
D = 0.059337, p-value = 0.3105
alternative hypothesis: two-sided
```

Anderson-Darling normality test

```
data:  sinan_hans_uf_2001_2022_rs$n
A = 0.86103, p-value = 0.02672
```

Já aqui temos resultados conflitantes, segundo os testes de Shapiro-Wilk e Anderson-Darling ambas as distribuições não seguem uma distribuição normal, ambos rejeitando a hipótese nula de uma distribuição normal, o histograma da distribuição de contagem de casos também não demonstra um comportamento normal. Mas o teste de Kolmogorov-Smirnov não rejeita a hipótese nula de uma distribuição normal, ou seja, indica que os dados seguem sim uma distribuição normal.

4.3 Comparando a prevalência de hanseníase entre estados

Agora vamos comparar a prevalência de hanseníase entre os estados brasileiros ao longo dos anos. Para isso, utilizaremos o teste de Mann-Whitney U para comparar as médias de duas amostras independentes. Vamos separar os dados por UF (Bahia e Rio Grande do Sul) e iremos comparar os valores de prevalência de hanseníase entre esses dois estados.

Wilcoxon rank sum test with continuity correction

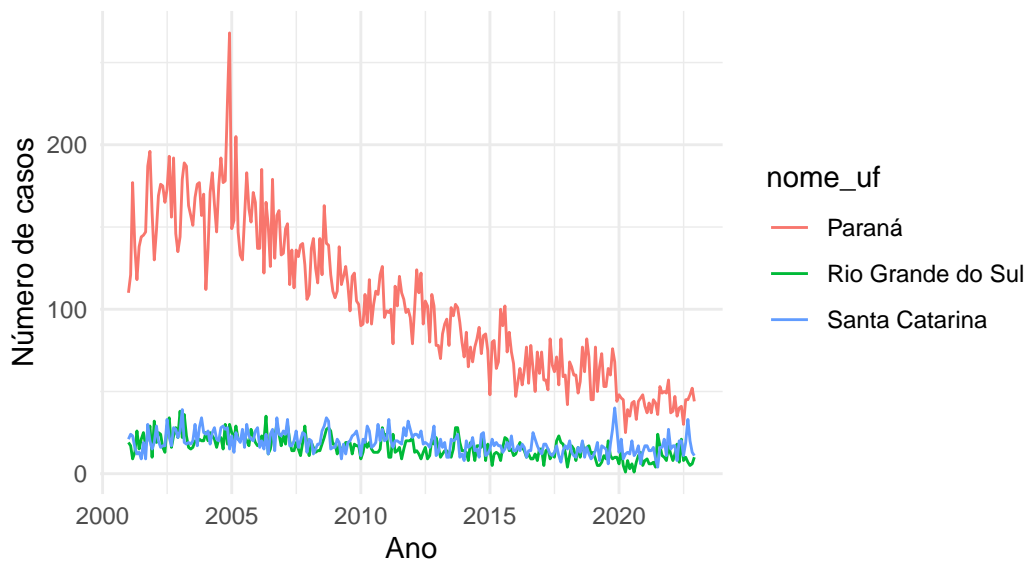
```
data:  sinan_hans_2001_2022_anomes_ba_rs_matrix$Bahia and sinan_hans_2001_2022_anomes_ba_rs_r
W = 69696, p-value < 0.00000000000000022
alternative hypothesis: true location shift is not equal to 0
```

O resultado do teste U de Mann-Whitney foi significativo ($p\text{-valor} < 0.05$), indicando que há uma diferença significativa na prevalência de hanseníase entre os estados da Bahia e do Rio Grande do Sul ao longo dos anos.

4.4 Comparando o número de casos de hanseníase em diferentes estados da mesma região

Agora vamos comparar o número de casos de hanseníase em uma região específica no período selecionado, usaremos os estados do Sul, portanto Paraná, Rio Grande do Sul e Santa Catarina, para observarmos se a diferença de casos também está presente, ou se os indicadores socioeconômicos influenciam a prevalência encontrada. Utilizaremos como um delineamento em blocos, onde os blocos serão as UFs, e os grupos (ou “tratamentos”) serão os anos/meses observados.

Distribuição de casos de Hanseníase no Sul do Brasil
Dados do SISVAN



Friedman rank sum test

```
data: sinan_hans_2001_2022_anomes_sul$n, sinan_hans_2001_2022_anomes_sul$ano_mes and sinan_l
Friedman chi-squared = 526.14, df = 263, p-value < 0.00000000000000022
```

Multiple comparisons between groups after Friedman test

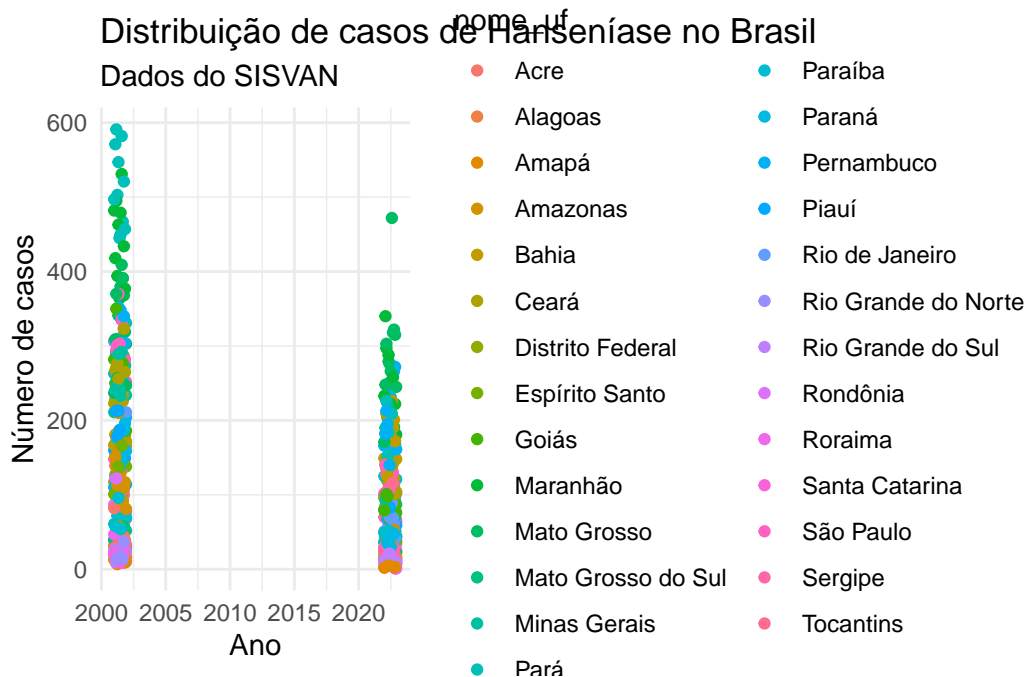
alpha: 0.05

Comparisons

	obs.dif	critical.dif	stat.signif
Paraná-Rio Grande do Sul	440	55.00947	TRUE
Paraná-Santa Catarina	352	55.00947	TRUE
Rio Grande do Sul-Santa Catarina	88	55.00947	TRUE

[illegible]

4.5 Comparando o número de casos entre os primeiros anos de notificação e os últimos



Friedman rank sum test

```
data:  sinan_hans_2001_2022_anomes_inicio_fim$n, sinan_hans_2001_2022_anomes_inicio_fim$nome
Friedman chi-squared = 582.9, df = 26, p-value < 0.00000000000000022
```

5 Conclusão

A hanseníase é uma doença infecciosa que ainda persiste em muitas partes do mundo, especialmente em países em desenvolvimento como o Brasil. Neste relatório, utilizamos testes estatísticos não paramétricos para avaliar as diferenças na prevalência de hanseníase nos estados do Brasil ao longo dos anos. Fizemos inicialmente uma comparação entre dois estados de regiões distintas, um do nordeste onde sabidamente existe maior prevalência da doença e outro de uma região menos afetada do sul e podemos observar através do teste U de Mann-Whitney que haviam diferenças entre os números de casos entre os estados. Além disso, comparamos o número de casos de hanseníase em uma região específica para observarmos se essa diferença de casos também está presente, ou se os indicadores socioeconômicos influenciam a prevalência encontrada. E mesmo com indicadores socioeconômicos mais próximos, os estados do Sul apresentam diferenças significativas no número de casos de hanseníase no período observado.

Ano Inicial	Ano Final	obs.dif	critical.dif	Significativo	P-valor
2001-01	2001-02	2.50	194.54	FALSE	1.00
2001-01	2001-03	85.50	194.54	FALSE	1.00
2001-01	2001-04	15.00	194.54	FALSE	1.00
2001-01	2001-05	74.00	194.54	FALSE	1.00
2001-01	2001-06	39.00	194.54	FALSE	1.00
2001-01	2001-07	31.50	194.54	FALSE	1.00
2001-01	2001-08	93.00	194.54	FALSE	1.00
2001-01	2001-09	51.50	194.54	FALSE	1.00
2001-01	2001-10	100.00	194.54	FALSE	1.00
2001-01	2001-11	19.00	194.54	FALSE	1.00
2001-01	2001-12	106.00	194.54	FALSE	1.00
2001-01	2022-01	317.00	194.54	TRUE	0.00
2001-01	2022-02	238.00	194.54	TRUE	0.00
2001-01	2022-03	177.50	194.54	FALSE	0.19
2001-01	2022-04	311.50	194.54	TRUE	0.00
2001-01	2022-05	177.50	194.54	FALSE	0.19
2001-01	2022-06	239.50	194.54	TRUE	0.00
2001-01	2022-07	292.50	194.54	TRUE	0.00
2001-01	2022-08	143.50	194.54	FALSE	1.00
2001-01	2022-09	202.50	194.54	TRUE	0.03
2001-01	2022-10	253.00	194.54	TRUE	0.00
2001-01	2022-11	270.00	194.54	TRUE	0.00
2001-01	2022-12	352.50	194.54	TRUE	0.00
2001-02	2001-03	83.00	194.54	FALSE	1.00
2001-02	2001-04	17.50	194.54	FALSE	1.00
2001-02	2001-05	71.50	194.54	FALSE	1.00
2001-02	2001-06	41.50	194.54	FALSE	1.00
2001-02	2001-07	34.00	194.54	FALSE	1.00
2001-02	2001-08	90.50	194.54	FALSE	1.00
2001-02	2001-09	54.00	194.54	FALSE	1.00
2001-02	2001-10	97.50	194.54	FALSE	1.00
2001-02	2001-11	16.50	194.54	FALSE	1.00
2001-02	2001-12	108.50	194.54	FALSE	1.00
2001-02	2022-01	319.50	194.54	TRUE	0.00
2001-02	2022-02	240.50	194.54	TRUE	0.00
2001-02	2022-03	180.00	194.54	FALSE	0.16
2001-02	2022-04	314.00	194.54	TRUE	0.00
2001-02	2022-05	180.00	194.54	FALSE	0.16
2001-02	2022-06	242.00	194.54	TRUE	0.00
2001-02	2022-07	295.00	194.54	TRUE	0.00
2001-02	2022-08	146.00	194.54	FALSE	1.00
2001-02	2022-09	205.00	194.54	TRUE	0.02
2001-02	2022-10	255.50	194.54	TRUE	0.00
2001-02	2022-11	272.50	194.54	TRUE	0.00
2001-02	2022-12	355.00	194.54	TRUE	0.00
2001-03	2001-04	100.50	194.54	FALSE	1.00
2001-03	2001-05	11.50	194.54	FALSE	1.00
2001-03	2001-06	124.50	194.54	FALSE	1.00
2001-03	2001-07	117.00	194.54	FALSE	1.00