任天堂股票分析报告

统计 1701 班

尹恒

U201710027

2020 年 4 月 16 日

目录

1 实验目的

本次实验选取从事电子游戏和玩具的开发、制造与发行的日本公司任天堂，并对其股票收益率进行分析

1. 求出每日收益率，作图将数据可视化，并分析结论。

2. 找一个收益率 (例如 1%)，将 (1%) 的收益率视为优秀，求优秀比例 95% 的置信区间。

3. 中位数检验：求出第一年收益率的中位数，用第二年收益率序列做比较，看是否相等。

4. 求收益率序列上下四分位数的置信区间。

5. 股票收益率序列有无上升趋势。

6. 选取另一只股票作相关性检验。

2 数据的可视化分析

本次实验选取任天堂股票作为分析，股票代码为 7974，在英为财情下载数据，并对 2018 年 4 月 11 日到 2020 年 4 月 11 进行分析。

2.1 数据的可视化

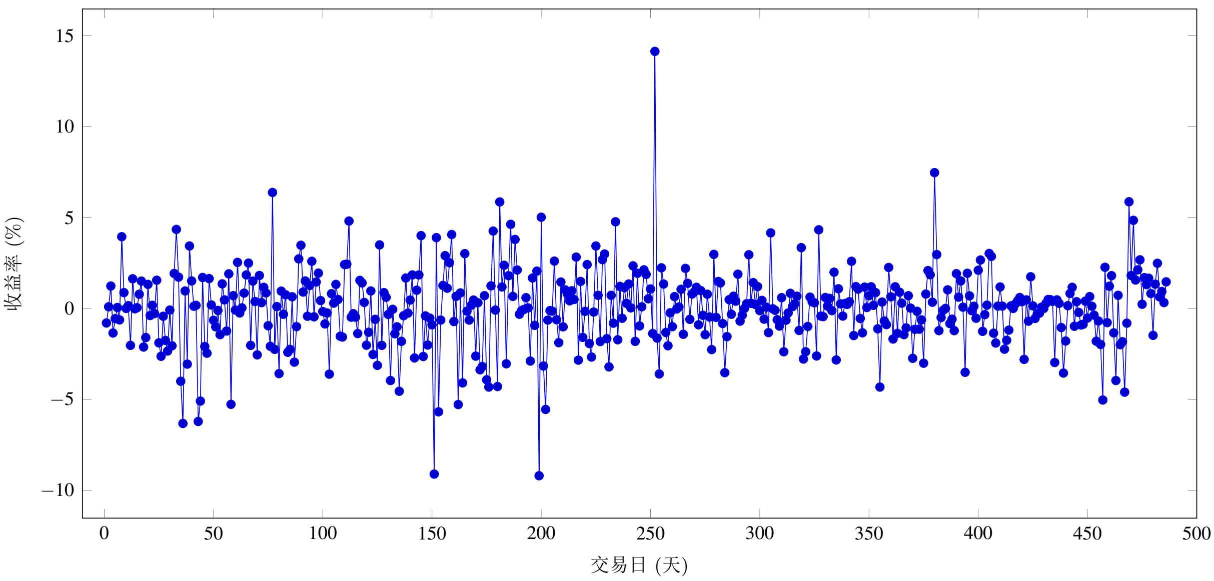

图 1: 2018.4-2020.4 任天堂股票收益率变化折线图

2.2 整体收益率变化分析

根据计算公式股票日收益率 =(今日收盘价-前日收盘价)/前日收盘价 ×100%. 全年收益率如上表。经计算该股票的收益率均值为 1.52%，方差为 0.000459，以上数据说明该<u>股票全年收益为正收益</u>，方差较小但极差很大，说明全年整体比较稳定但是偶尔会出现很大的波动。

3 基于近似正太分布的置信区间估计

3.1 实验原理

数据 察看含有 n 个独立基本实验观测值的样本，并记 Y 为指定事件发生次数

假定条件

1. n 次基本实验互相独立
2. 指定事件发生的概率 p 是常数

方法 A 对于 $n <= 30$, 利用表，得到精确值

方法 B 对于 $n > 30$，或置信系数没有在表中列出的，用下列正态分布逼近

$$L = \frac{Y}{n} - z_{1-\alpha/2}\sqrt{\frac{Y(n-Y)}{n^3}}$$

和

$$U = \frac{Y}{n} + z_{1-\alpha/2}\sqrt{\frac{Y(n-Y)}{n^3}}$$

3.2 优秀比例的置信区间

	N	最大值	最小值	均值	标准偏差
收益率	486	14.12%	-9.19%	0.01971%	0.021429434

表 1: 任天堂股票收益率统计描述

观察表 1，可得股票收益率分布在-9.19% 到 14.12% 之间，我们可以设定大于 1.5% 为优秀，经过 Excel 筛选得样本中有 91 天为优秀，则样本优秀率为：

$$\hat{p} = \frac{91}{486} = 0.18724$$

接下来计算优秀比例的 95% 的置信区间。本次实验显然为大样本情况，所以使用方法 B 可以求出置信上下限：

$$L = \frac{Y}{n} - z_{1-\alpha/2}\sqrt{\frac{Y(n-Y)}{n^3}} = \frac{91}{486} - 1.96\sqrt{\frac{91(486-91)}{486^3}} = 0.15255945850318617$$

和

$$U = \frac{Y}{n} + z_{1-\alpha/2}\sqrt{\frac{Y(n-Y)}{n^3}} = \frac{91}{486} + 1.96\sqrt{\frac{91(486-91)}{486^3}} = 0.22192613820463278$$

所以<u>优秀比例 p 的 95% 的置信区间为 [0.152559,0.221926]</u>。

4 中位数检验

4.1 实验原理

数据 令 $X_1,X_2,...,X_n$ 是一组随机样本，数据由 X_i 的观测值组成。

假定条件

1. 这些 X_i 是随机样本 (即, 它们是独立同分布的随机变量).
2. 度量尺度至少是顺序的.

检验统计量 在这个检验中我们将用两个检验统计量，令 T_1 等于观测值中小于等于 x^* 的个数，T_2 等于观测值中小于 x^* 的个数，那么，当数据中没有数严格等于 x^* 的数时，T_1=T_2, 否则，T_1 大于 T_2。

零分布 检验统计量 T_1 和 T_2 的零分布是二项分布，参数 $n=$ 样本量，$p=p*$ 和零假设一样对于其他 n,p 值，用正态分布逼近，即，T 的近似分位数 x_q 为

$$x_q = n\times p + z_q\sqrt{n\times p\times(1-p)}$$

其中,z_q 是标准正态随机变量的 q 分位数。

假设 令 x^*,p^* 为指定的值，$0<p^*<1$, 则假设可能是如下三种形式中的一种。

A.(双边检验)

H_0：第 p^* 个总体的分位数为 x^*[这等价于 $H_0:P(X\le x^*)\ge p^*$ 且 $P(X<x^*)\le p^*$]

H_1：x^* 不是第 p^* 个总体的分位数

对于给定的 n 和 p^*，$Y\sim B(n,p^*)$, 查表 $P(Y\le t_1)=\alpha_1\approx\frac{\alpha}{2},P(Y\le t_2)=1-\alpha_2,\alpha_1+\alpha_2\approx\alpha$

故，拒绝域 $\{T_1\le t_1\}\cup\{T_2>t_2\}$，显著性水平为 $\alpha_1+\alpha_2$。

B.(左边检验)

$H_0: x_{p^*}\le x^*$(这等价于 $P(X\le x^*)\ge p^*$)

$H_1: x_{p^*}>x^*$

拒绝域：T_1 太小，查表 $P(Y\le t_1)=\alpha.[T_1\le t_1]$

C.(右边检验)

$H_0: x_{p^*}\ge x^*$(这等价于 $P(X<x^*)\le p^*$)

$H_1: x_{p^*}<x^*$

拒绝域：T_2 太大，查表 $P(Y\le t_2)=1-\alpha.[T_2>t_2]$

4.2 中位数检验

为了更好的观察和对比两年的收益率，分别计算出每年的四分位数表如下

	N	最小值	25% 分位数	中位数	75% 分位数	最大值
2018-2019	246	-9.19%	-1.54%	0.03%	1.3425%	6.37%
2019-2020	240	-5.03%	-0.86%	0.145%	0.9675%	14.12%

表 2: 任天堂股票收益率的四分位数

根据上表绘制出每年的收益率箱型图如下

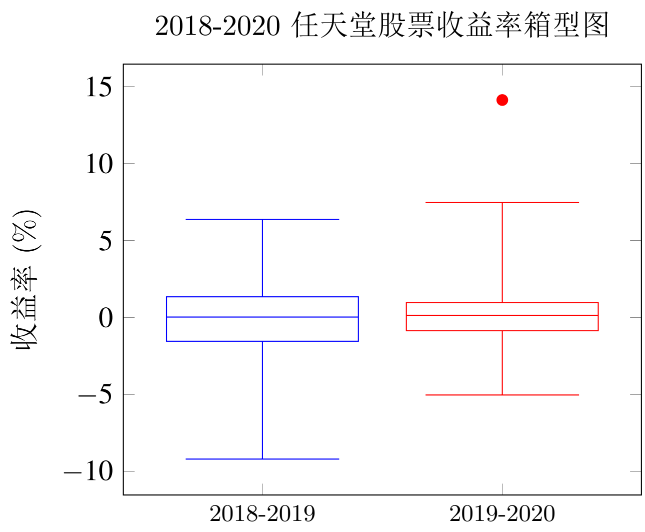

图 2: 任天堂股票收益率年度箱型图

通过分析图 2，可以发现，两年的中位数和上下四分位数基本一致，2019-2020 出现了一个偏离程度较大的值，剔除后两年的箱型图基本一致，所以任天堂的股票收益率两年变化不大。

现在对收益率的中位数进行定量分析，用第一年 (2018) 的中位数对第二年 (2019) 的中位数进行假设检验，即检验:

H_0:2019 年收益率的中位数是 0.003

H_1:2019 年收益率的中位数不是 0.003

选取检验统计量 T_1 是观测值中小于等于 x^* 的个数;T_2 是观测值中严格小于 x^* 的个数，经过筛选与计数，$T_1 = 121, T_2 = 107$。

对于大样本，可以利用正态分布进行逼近，令置信水平 $\alpha = 5\%$，可以求出拒绝域的边界：

$T_l = np - z_{\frac{\alpha}{2}}\sqrt{np(1-p)} = 240\times 0.5 - 1.96\times\sqrt{240\times 0.5\times(1-0.5)} = 104.82$

$T_r = np + z_{\frac{\alpha}{2}}\sqrt{np(1-p)} = 240\times 0.5 + 1.96\times\sqrt{240\times 0.5\times(1-0.5)} = 135.18$

故拒绝域为 $\{T_1 \le 104.82\}\cup\{T_2 \ge 135.18\}$, 由于 $T_2 = 107$ 不在拒绝域中，故接受原假设，即<u>任天堂股票 2019 年收益率的中位数约等于 0.0003</u>.(该结论与肉眼观察箱型图得出结论一致)。

5 分位数的置信区间

5.1 实验原理

数据 数据由独立同分布的随机变量 $X_1,X_2,...,X_n$ 的观测组成，$X^{(1)} \leq X^{(2)} \leq ... \leq X^{(r)} \leq ... \leq X^{(s)} \leq ... \leq X^{(n)}$ 为次序统计量,$1 \leq r < s \leq n$. 希望找到 p^*(未知数) 分位数

假定条件

1. 这些 X_i 是随机样本.
2. X_i 的度量尺度至少是顺序的.

双边检验的置信区间为:

$$P(X^{(r)} \leq x_{p^*} \leq X^{(s)}) \geq 1-\alpha_1-\alpha_2$$

单边置信区间为：

$$P(X^{(r)} \leq x_{p^*}) \geq 1-\alpha_1$$

$$P(x_{p^*} \leq X^{(s)}) \geq 1-\alpha_2$$

方法 A(**小样本**) 查表 3，$p=p^*,n=$ 样本容量, 找到对应于 $1-\alpha_1$ 的 $s-1$, 和对应于 $1-\alpha_2$ 的 $r-1$ 即可.

方法 B(**大样本近似**) 对于 n 大于 20，可以用基于中心极限定理的逼近，计算

$$r^* = np^* + z_{\alpha/2}\sqrt{np^*(1-p^*)}$$

$$s^* = np^* + z_{1-\alpha/2}\sqrt{np^*(1-p^*)}$$

所得 r^* 和 s^* 向上取整.

5.2 收益率上下四分位数检验

本实验显然为大样本情况，利用中心极限定理正态逼近得到公式:

$$r^* = np^* + z_{\alpha/2}\sqrt{np^*(1-p^*)}$$

$$s^* = np^* + z_{1-\alpha/2}\sqrt{np^*(1-p^*)}$$

上四分位点: 代入 $p^* = 0.25, \alpha = 0.05$

$$r^* = 486 \times 0.25 - 1.96 \times \sqrt{486 \times 0.25 \times (1-0.25)} = 102.78995456980395$$

$$s^* = 486 \times 0.25 + 1.96 \times \sqrt{486 \times 0.25 \times (1-0.25)} = 140.21004543019603$$

故取 $r=103, s=141$, 将收益率从小到大排序后可得:

$$X^{(103)} = -0.0137, X^{(141)} = -0.0081$$

所以可得收益率上四分位数的 95% 置信区间为

$$[-0.0137, -0.0081]$$

下四分位点: 代入 $p^* = 0.75, \alpha = 0.05$

$$r^* = 486 \times 0.75 - 1.96 \times \sqrt{486 \times 0.75 \times (1-0.75)} = 345.78995456980397$$

$$s^* = 486 \times 0.75 + 1.96 \times \sqrt{486 \times 0.75 \times (1-0.75)} = 383.21004543019603$$

故取 $r = 103, s = 141$, 将收益率从小到大排序后可得:

$$X^{(346)} = 0.009, X^{(384)} = 0.0134$$

所以可得收益率下四分位数的 95% 置信区间为 $[0.009, 0.0134]$.

6 符号检验

6.1 实验原理

数据 数据由随机变量序列 $X_1, X_2, ..., X_n$ 的观测值组成，把随机变量进行配对分组 (X_1, X_{1+c}), $(X_2, X_{2+c}), ..., (X_{n'-c}, X_{n'})$, 如果 n' 是偶数，则 $c = n'/2$, 如果 n' 是奇数，则 $c = (n'+1)/2$, 并且除去中间的随机变量, 如果 $X_i < X_{i+c}$, 则用“+”替代 $X_i < X_{i+c}$，如果 $X_i > X_{i+c}$, 则用”-” 代替 $X_i > X_{i+c}$.

假定条件

1. 随机变量 $X_1, X_2, ..., X_n$ 是相互独立的.

2.X_i 的度量尺度至少是顺序的.

3.X_i 是同分布或有某种趋势; 即后面的随机变量更可能比前面的大 (或反之亦然).

检验统计量 $T =' +'$ 的个数

假设 对于大样本，可用正态逼近 $X \sim B(n, \frac{1}{2}) \sim N(\frac{n}{2}, \frac{n}{4})$

A.(双边检验)

H_0:$P(+) = P(-)$(没有出现趋势),H_1:$P(+) \neq P(-)$

故，拒绝域 $\{T \le t\} \cup \{T \ge n-t\}$,

$$t \approx \frac{1}{2}(n + z_{\frac{\alpha}{2}}\sqrt{n})$$

$p-$ 值 $= 2min\{P(T \le t_{obs}), P(T \ge t_{obs})\}$

B.(左边检验)

H_0:$P(+) \geq P(-)$(没有下降趋势),H_1:$P(+) < P(-)$

故，拒绝域 $\{T \leq t\}$,

$$t \approx \frac{1}{2}(n + z_\alpha \sqrt{n})$$

$p-$ 值 $= P(T \leq t_{obs})$

C.(右边检验)

H_0:$P(+) \leq P(-)$(没有上升趋势),H_1:$P(+) > P(-)$

故，拒绝域 $\{T \geq n - t\}$,

$$t \approx \frac{1}{2}(n + z_\alpha \sqrt{n})$$

$p-$ 值 $= P(T \geq t_{obs})$

6.2 趋势性检验

不妨设原假设和备择假设分别为：$H_0 =$ 收益率没有上升趋势,$H_1 =$ 收益率有上升趋势, 利用 Cox 和 Stuart 趋势性检验方法, 利用 Excel 的函数功能将数据两两配对，数据共有 486 条，所以将数据配对为 $(X_1, X_{1+243}), (X_2, X_{2+243}), ..., (X_{486-243}, X_{486})$, 共 243 组, 没有节点, 若 $X_i < X_{i+243}$, 则记为"+". 用 excel 筛选出 $T = "+"$ 的个数为 121, 显然为大样本所以用正态逼近,T 越大越拒绝原假设,

$$t = 243 \times \frac{1}{2} + 1.6449 \times \sqrt{243 \times \frac{1}{2} \times \frac{1}{2}} = 134.3207$$

所以拒绝域为 $\{T >= 135\}$,$T = 121$ 不在拒绝域中，所以接受原假设，即任天堂股票收益率没有上升趋势。

6.3 相关性检验

本文选取与任天堂同类的游戏发行公司万代南梦宫的股票与任天堂的股票作相关性检验, 获取相同时间段万代南宫梦的股票收益率，将其作为 Y，与 X(任天堂股票收益率) 进行相关性检验。设：

H_0 : 任天堂股票收益率与万代南宫梦股票收益率没有正相关关系

H_1 : 任天堂股票收益率与万代南宫梦股票收益率有正相关关系

用 excel 将同一天的两只股票放在同一行，进行配对，将配对好的数据按照 X(任天堂股票收益率) 升序排列，将排列之后的数据 Y 按照趋势性检验中的步骤进行配对，共 243 组，筛选出 $T = "+"$ 的个数为 117，显然为大样本所以用正态逼近,T 越大越拒绝原假设,

$$t = 243 \times \frac{1}{2} + 1.6449 \times \sqrt{243 \times \frac{1}{2} \times \frac{1}{2}} = 134.3207$$

所以拒绝域为 $\{T >= 135\}$,$T = 117$ 不在拒绝域中，所以接受原假设，即任天堂股票收益率和万代南梦宫股票收益率没有正相关关系。

7 实验总结

本次实验中我学会了用 SPSS 和 excel 两种统计分析软件去分析问题，用 Latex 排版，通过本次实验我对数据的可视化，各种图表，比例置信区间，分位数检验和符号检验有了更深入的学习和了解，用书本知识去解决现实问题。