

МС-9. Описательная статистика

1. Следующие 40 чисел – это умноженные на 10000 и округленные до ближайшего целого дневные логарифмические доходности акции компании ABC: -208, 94, -739, -58, 179, 78, -237, 23, 236, 277, 301, 241, -165, 67, 211, -369, 666, 84, 220, -83, 166, -496, 205, 108, 45, 197, -452, -268, 150, 27, -206, -7, -263, -7, -611, 556, -620, 302, 403, -289. Будем называть их преобразованными доходностями (ПД). Финансовый аналитик Глеб предполагает, что преобразованные доходности (как и исходные) приближенно распределены по нормальному закону. Чтобы проверить свое предположение Глеб нашел нижнюю квартиль L и верхнюю квартиль H нормального распределения $N(\mu, \sigma^2)$, для которого μ – это среднее арифметическое ПД, а σ – эмпирическое стандартное отклонение ПД. Затем Глеб подсчитал количество ПД, попавших в интервал от L до H (надеясь, что в этот интервал попадет половина ПД). Результат этого вычисления показался ему недостаточно убедительным. Чтобы окончательно развеять сомнения относительно нормальности ПД, Глеб построил на одном рисунке графики функций: $\hat{F}(x)$ и $F(x)$, где $\hat{F}(x)$ – эмпирическая функция распределения ПД, а $F(x)$ – функция распределения $N(\mu, \sigma^2)$. В качестве меры совпадения двух графиков Глеб решил использовать расстояние d между функциями $\hat{F}(x)$ и $F(x)$, которое он вычислил, исходя из определения: $d = \sup |\hat{F}(x) - F(x)|$. В ответе укажите результаты вычислений Глеба: 1) среднее арифметическое ПД; 2) эмпирическое стандартное отклонение ПД; 3) квартили L и H ; 4) количество ПД, попавших в интервал от L до H ; 5) расстояние между функциями $\hat{F}(x)$ и $F(x)$.

Ответ:

-6,05 - среднее арифметическое ПД

311,3476 - эмпирическое стандартное отклонение ПД

-216,050786 - квартиль L

203,950786 - квартиль H

19 - количество ПД, попавших от L до H

0,12378 - расстояние между функциями распределений

2. Форматирование данных:

Независимые наблюдения нормально распределенной случайной величины X , описывающей доходность определенного вида фьючерсов, представлены в виде неупорядоченного вариационного ряда:

{196.92971; NA; 136.07691; 196.34805; 199.53462; 170.20138; 197.97725; 202.0997; NA; 194.73865; 168.11473; 137.27056; 227.51421; 187.43231; 135.74843; 168.04959; 158.53354; 186.88781; 207.99242; 174.9859; 200.20625; NA; 224.08094; 170.5565; 158.7361; 180.9028;

182.55506; 221.7838; NA; NA; NA; 239.06614; 158.3426; 195.5757; 164.48166; 166.60759; 247.77945; 190.90885; NA; 339.985965; 174.58393; NA; 196.03865; NA; 203.91763; 170.6796; NA; 223.15341; 222.0477; 208.74949; 212.50459; 198.35273; NA; 152.41688; NA; 369.864755; 230.896; 182.36959; 153.89932; 240.00281; NA; 217.4074; 219.50106; 170.24959; NA; 184.27259; 178.06786; 186.45008; NA; 171.8145; 200.80527; 201.05794; 183.38469; 173.17188; 129.8169; 182.97317; 180.65213; 214.35604; 198.06513; 195.32153; 186.26798; NA; NA; NA; NA; 219.56085; 195.18918; 184.03913; 171.1902; 200.14447; 192.77745; 179.95669; 177.76787; 193.05699; 216.59228; 192.21672; 202.00829; 196.74733; 169.85218; 198.18508; NA; 213.06066; 205.42989; 159.57486; 197.35776; 168.14253; 197.65186; NA; NA; 206.6751; 142.65346; 200.02902; 197.99453; 182.79758; NA; 199.8229; 215.48279; 185.62296; 157.46312; 229.69705; 210.81753; 200.34913; 192.05463; 71.07685500000001; 157.48724; NA; 178.2801; 179.3978; 181.28785; 183.23624; 197.99604; 190.79779; 174.78006; 176.66681; 188.99502; 166.40737; 168.10608; 174.66107; 166.65371; NA; 195.34543; 194.31794; 150.14698; 218.22001; 219.45564; 230.20715; 165.10735; 207.49599; 182.98615; 194.35029; 154.62618; 150.79688; 195.57284; 201.92714; 203.45877; 172.0433; NA; 159.3953; 176.32943; 173.94611; 219.23781; 183.56996; 193.17269; 187.71603; 207.17507; 184.69483; 155.43353; 191.20643; 179.8024; 195.10684; 234.1981; 231.39562; NA; NA; 187.65472; NA; 225.66359; NA; 180.61912; 209.89129; NA; 190.70637; 180.91974; 216.39233; 225.06374; 165.82589; 177.65018; NA; 182.17528; 191.46474; 220.61831; 163.68818; 187.51613; 195.09282; 175.87486; 186.19105; 149.79936; 198.84328; 152.91778; 223.71218; 176.80615; 207.54141; 195.86928; 211.87501; 177.74898; 223.8982; 208.12129; 213.85756; 156.97355; 178.9495; 156.48682; 184.85755; 202.30561; 217.3244; 189.23265; 215.32331; NA; 138.64001; NA; 179.31733; 159.37454; 215.14106; 187.80021; 195.50795; 181.82473; 183.6953; 198.15892; NA; 202.55826; NA; 176.20312; 173.75703; 141.42246; 206.44492; 172.39673; 177.01227; 219.86468; NA; 181.23446; 156.2762; NA; 100.955645; 163.63176; 164.60035; 200.64875; 219.81358; 207.8252; 193.1447; 224.34229; 176.0246; 197.2936; 224.94001; 191.90226; NA; 206.45662; 197.52237; 164.21773; 167.85524; 244.50199; 194.69737; 190.95449; 185.85411; NA; 231.7228; 185.34329; 207.02483; 205.83597; NA; 212.50914; NA; 227.26332; 216.09838; 180.63413; 157.3706; 200.64688; NA; 205.0958; 197.44369; 195.75397; 185.39637; 212.56476; 221.23139; 223.00506; NA; 227.41976; 195.25196; 204.67768; 191.34718; 182.85024; 161.94313; 188.26786; NA; 170.67867; 151.12892; 196.39718; 179.52197; NA; NA; 167.24265; 182.42814}.

Очистите ряд от пропусков и вылетов. Получите точечные оценки описательной статистики двумя способами:

- автоматизированными формулами в ячейках MS Excel (используя функции Python);
- вызовом описательной статистики.

Самостоятельное задание

1. Форматирование данных:

Независимые наблюдения нормально распределенной случайной величины X , описывающей приращение стоимости акций некоторой компании, представлены в виде выборки:

{-268.245074625; -75.409445; -164.047126; -127.629816; -171.590449; -176.739644; -185.531303; -120.60225; -115.617363; -155.691373; -124.245553; -185.984563; -214.001891; -126.207037; -110.357723; NA; -134.379138; -171.282078; -137.720122; -131.077658; -124.196612; -151.227615; -150.862058; -153.82287; -163.667714; NA; NA; -187.811033; -177.475979; -187.550272; -142.17284; -182.366075; -182.12927; -156.468605; -175.389551; -149.055159; -

189.424868; -129.56455; -140.602316; -115.86738; -155.7073; NA; NA; -166.99668; -127.365801; -135.619692; -131.423629; -141.984776; -172.295086; -125.943666; -126.736385; -189.901613; -201.366008; -105.687576; -152.649723; -211.140098; -127.665601; -160.893549; -148.472221; -160.252389; NA; -152.406848; -153.706815; -308.698052875; -115.237311; NA; NA; -134.917789; -103.491199; -132.680416; -131.872446; -180.504646; -199.339202; -175.526342; -176.675147; -187.376153; -190.623923; -144.716348; -155.256536; -170.094714; -123.945107; -159.080007; -136.847111; NA; -210.088116; NA; -105.120859; -135.334757; -77.295984; -169.148997; NA; NA; -129.446608; -156.522821; -143.183933; -180.604798; NA; -138.248728; -193.280969; NA; -158.890902; NA; -115.927408; -141.311527; -145.078179; -160.788705; -132.037863; -160.436474; -135.113808; -154.127932; -108.828174; NA; -152.299651; -128.705452; -152.049001; NA; -182.279822; -142.537416; -144.660644; -134.21196; -207.085892; -127.030172; -153.099842; -147.900537; -176.899247; -206.549914; -81.084677; -160.717528; -169.421077; -188.082487; -138.340435; -111.001113; -152.615313; -178.398956; -146.958429; -210.814119; -161.057824; -137.728629; NA; -140.306312; NA; -182.446721; -134.650344; -137.690554; -144.332193; -168.488586; -144.876708; -165.50784; -177.693515; -126.407513; -165.451324; -213.521504; -160.687058; -213.714958; -177.258508; NA; -165.569893; NA; NA; -166.715938; -152.18066; -123.243822; -150.532819; -153.903795; NA; NA; NA; -272.17513625; -166.443929; -168.655442; -141.983367; -173.784691; -137.910001; -148.222066; -168.314316; -197.648146; NA; -103.420587; NA; -186.009116; -153.630412; -223.89061; -130.474652; -170.494017; -169.003448; -214.645533; -149.964583; -183.661943; -170.656833; -163.506458; -135.104197; -147.652628; -108.906857; NA; NA; -142.998136; -148.360945; -199.256758; -120.30029; -91.633162; -175.595035; -122.617492; -152.726522; NA; -159.690058; -111.624624; -144.387548; -149.927723; -127.957205; -191.58146; -132.586363; -150.622; -148.182417; -89.330805; -190.955068; -149.932124; NA; -210.78763; -118.958718; -135.848819; -161.366843; -145.930802; NA; -184.262147; -143.29328; -130.249766; -174.656404; -93.20679; -141.520405; -169.723387; -124.365752; -154.650156; -218.207332; -153.255727; -169.10628; -88.522842; -169.018656; -100.029771; -200.398643; -148.085069; -91.984848; NA; -123.204525; -134.503908; -99.836579; -140.370375; -169.123292; -183.407501; -201.081003; -126.584147; -155.042613; -180.587576; -93.104521; NA; -143.096813; -141.501224; -128.745108; NA; -203.430507; -144.629077; -128.677943; -145.82055; -123.802731; NA; -86.250199; -156.997474; -195.589801; -178.318793; -132.050742; NA; -162.895506; -147.38235; NA; -173.935485; -99.81084; -153.058583; -116.573012; -151.26499; -130.606171; -121.97044; -127.623157; -165.410891; -175.517653; NA; -112.097073; NA; NA; -157.93647; -114.348722; -161.607122; -161.409723; -167.229725; NA; -159.918433; -148.405462; -128.422548; -147.9132; -217.696706; -158.670811; -164.154015; -151.094936; NA; -103.798141; -123.517802; -181.351361; -189.887351; -143.519768; -177.075376; NA; -102.871813}.

Очистите ряд от пропусков и вылетов. Получите точечные оценки описательной статистики двумя способами:

- автоматизированными формулами в ячейках MS Excel;
- вызовом описательной статистики.

Домашнее задание

1. Форматирование данных:

Независимые наблюдения нормально распределенной случайной величины X , описывающей приращение стоимости типового контракта страховой фирмы, представлены в виде неупорядоченного вариационного ряда:

{-194.3779; -177.7017; NA; -190.9091; -170.5325; -168.1535; -197.1981; -122.5442; -181.6037; -196.2927; NA; -87.3432; -207.3771; -143.9177; -189.7354; -194.9819; -165.5695; -165.3043; -171.8495; -168.1425; -169.1902; -151.1296; -208.1701; -169.2235; -189.7955; NA; -172.3725; -183.0272; -221.2799; -139.1868; -160.088; -124.0217; -117.0158; -196.7869; -168.8747; -199.7569; -170.2329; -157.5982; -182.6017; NA; -198.2757; -173.3341; -197.0221; -176.617; -174.4793; -183.0195; -137.0551; -166.65; -161.1568; -204.9911; -137.6051; -85.022; -174.6369; -156.7334; -154.6654; -181.0034; -165.9679; -185.8001; -164.5037; -191.5906; -164.718; -205.4555; -160.3067; -212.2588; -161.5747; -189.3608; -146.8366; -194.0602; -166.7969; -191.3017; -162.0666; -168.4299; -210.1714; NA; -154.0824; -168.8861; -181.1642; -167.2657; -188.7629; -143.8932; -184.1726; -148.415; -163.2136; -179.3092; -158.3453; -152.927; -169.3517; -144.0586; -166.0867; -157.7147; -159.9157; -147.8518; -175.8872; -124.973; -149.9945; -255.031825; -130.5202; -209.4644; -125.7248; -165.9567; NA; -163.1068; -157.7618; -177.4579; -193.1898; -172.972; -172.5896; -145.9725; -164.5577; -124.0417; -160.3465; -134.8341; -161.6984; -165.9788; NA; -143.0549; -132.854; -137.7801; -193.0964; -184.9669; -151.5307; -185.057; -204.1557; -170.7604; -162.6274; -170.9198; -139.1593; -187.4823; -223.7508; -189.2962; -164.1744; -162.9236; -146.3165; NA; -163.4765; -163.3982; -163.5342; -198.4725; -202.5708; -178.5649; -224.0032; -168.6664; -195.6848; -158.2784; -132.0059; -162.5579; -143.5323; -175.7604; -177.9567; -151.2079; -125.5979; -187.1473; NA; -175.7593; -151.24; NA; -202.8399; -185.4153; -168.6321; -181.5665; -205.8899; -277.0064; NA; -173.3101; -128.5103; -172.5604; -154.2079; NA; -194.478; -161.6489; -163.2809; -176.994; -165.0078; -162.7372; -169.5874; -144.1649; -167.4255; -176.5829; -150.7951; -180.9425; -165.5115; -164.5462; -143.4277; -166.3309; -206.4914; -166.7015; -194.746; -148.9638; -191.2921; -169.9044; -186.2946; -215.7937; -149.0729; -146.4595; -189.1966; -165.4463; -212.4083; -186.8488; -165.2698; -146.8094; -185.0691; -159.2846; -124.6157; NA; -183.5672; -198.9645; -131.4182; NA; -196.0557; -202.8279; -164.7111; -140.1387; -184.423; -176.7473; -214.8683; -217.3961; -175.5419; -172.1291; -197.2019; -189.5526; -206.0636; -168.6153; -157.2746; -136.4666; -184.4911; -176.362; -161.5741; -177.9463; -172.3879; -172.095; -178.5666; -152.9071; NA; -176.4159; -162.2979; -161.6216; -182.6296; -173.2591; NA; -200.2; -131.9886; -173.4903; -174.1179; -143.78; -192.1052; NA; -155.9192; -156.6437; -157.2762; -144.4429; -147.86; NA; -177.8253; -175.6016; NA; -190.1326; -171.2815; -184.9209; -137.929; -164.9219; -175.851; -155.4942; -168.2686; -162.7573; -127.2019; -188.4408; -189.451; -189.4407; -161.4335; -219.0344; NA; -172.7585; -158.4946; -150.4887; -132.1469; -162.99; -169.8512; -162.7875; NA; -130.4107; -120.2871; -171.8417; -180.7133; -176.3202; -171.8288; -163.8611; -164.0547; -131.1598; -134.2916; -177.699; -268.3541; -154.5068; -167.0612; -136.12; -187.417; -170.5064; -134.3572; -179.4004; -187.3903; -203.872}.

Скопируйте и преобразуйте в столбец "А" данные вариационного ряда на лист "Лист3" Excel-файла (Python с использованием docx) и очистите ряд от пропусков и вылетов. Получите точечные оценки описательной статистики двумя способами:

- автоматизированными формулами в ячейках (используя функции Python);
- вызовом описательной статистики.

2. Следующие 36 чисел – это умноженные на 10000 и округленные до ближайшего целого дневные логарифмические доходности акции компании ABC: 295, -158, -51, 34, 109, -53, -185, 233, -251, -320, -274, -116, 370, -83, 149, 66, -108, -236, -87, -30, 113, -35, 56, -220, 20, -339, 90, 240, 100, -115, -137, 129, 99, -102, 299, -145. Будем называть их преобразованными доходностями (ПД). Финансовый аналитик Глеб предполагает, что преобразованные доходности (как и исходные)

приближенно распределены по нормальному закону. Чтобы проверить свое предположение Глеб нашел нижнюю квартиль L и верхнюю квартиль H нормального распределения $N(\mu, \sigma^2)$, для которого μ – это среднее арифметическое ПД, а σ – эмпирическое стандартное отклонение ПД. Затем Глеб подсчитал количество ПД, попавших в интервал от L до H (надеясь, что в этот интервал попадет половина ПД). Результат этого вычисления показался ему недостаточно убедительным. Чтобы окончательно развеять сомнения относительно нормальности ПД, Глеб построил на одном рисунке графики функций: $\hat{F}(x)$ и $N(\mu, \sigma^2)$, где $\hat{F}(x)$ – эмпирическая функция распределения ПД, а $F(x)$ – функция распределения $N(\mu, \sigma^2)$. В качестве меры совпадения двух графиков Глеб решил использовать расстояние d между функциями $\hat{F}(x)$ и $F(x)$, которое он вычислил, исходя из определения: $d = \sup |\hat{F}(x) - F(x)|$. В ответе укажите результаты вычислений Глеба: 1) среднее арифметическое ПД; 2) эмпирическое стандартное отклонение ПД; 3) квартили L и H ; 4) количество ПД, попавших в интервал от L до H ; 5) расстояние между функциями $\hat{F}(x)$ и $F(x)$.

Ответ:

-17,8611 - среднее арифметическое ПД

178,2157 - эмпирическое стандартное отклонение ПД

-138,0658 - квартиль L

102,3436 - квартиль H

18 - количество ПД, попавших от L до H

0,0871 - расстояние между функциями распределений

Help:

<https://en.wikipedia.org/wiki/Quantile>