STA2002 Assignment 4

Chen Ang (118010009)

1

(a)

$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$	$X_{(11)}$	$X_{(12)}$	$X_{(13)}$	$X_{(14)}$
5.2	5.5	6.3	7.1	7.5	8.7	9.2	9.8	10.5	10.9	11.1	11.8	12.7	14.4

$$\begin{aligned}\hat{\pi}_{0.25} &= \hat{\pi}_{3.75/15} \\ &= 0.25 \cdot \hat{\pi}_{3/15} + 0.75 \cdot \hat{\pi}_{4/15} \\ &= 0.25 \cdot X_{(3)} + 0.75 \cdot X_{(4)} \\ &= 6.9\end{aligned}$$

$$\begin{aligned}\hat{\pi}_{0.35} &= \hat{\pi}_{5.25/15} \\ &= 0.75 \cdot \hat{\pi}_{5/15} + 0.25 \cdot \hat{\pi}_{6/15} \\ &= 0.75 \cdot X_{(5)} + 0.25 \cdot X_{(6)} \\ &= 7.8\end{aligned}$$

$$\begin{aligned}\hat{m} &= \hat{\pi}_{7.5/15} \\ &= 0.5 \cdot \hat{\pi}_{7/15} + 0.5 \cdot \hat{\pi}_{8/15} \\ &= 0.5 \cdot X_{(7)} + 0.5 \cdot X_{(8)} \\ &= 9.5\end{aligned}$$

$$\begin{aligned}\hat{\pi}_{0.75} &= \hat{\pi}_{11.25/15} \\ &= 0.75 \cdot \hat{\pi}_{11/15} + 0.25 \cdot \hat{\pi}_{12/15} \\ &= 0.75 \cdot X_{(11)} + 0.25 \cdot X_{(12)} \\ &= 11.3\end{aligned}$$

(b)

Recall from Lecture Note 13-4

$$P\left(X_{(i)} < \pi_p < X_{(j)}\right) = \sum_{k=i}^{j-1} \binom{14}{k} p^k (1-p)^{14-k}$$

i.

$$i = 1, j = 5, p = 0.25 \implies P = 0.724$$

ii.

$$i = 3, j = 8, p = 0.35 \implies P = 0.841$$

iii.

$$i = 5, j = 10, p = 0.5 \implies P = 0.820$$

iv.

$$i = 8, j = 13, p = 0.75 \implies P = 0.861$$

(c)

i.

We have the test statistic

$$N^{-} := \sum_{i=1}^{14} 1_{(X_i < 8)} \overset{H_0}{\sim} \mathrm{Bin}(14, 0.5)$$

with realized value

$$n^{-} = 5$$

So the p-value is

$$P(N^{-} \leq 5; H_0) = \sum_{n=0}^{5} \binom{14}{n} (0.5)^{14} = 0.212 > \alpha$$

So we fail to reject H_0.

ii.

Define

$$W := \sum_{k=1}^{14} R_k^{\pm}$$

where $R_k^{\pm} = \mathrm{sign}(X_k - 8) \cdot R_k$ is the k-th signed rank.

"x"	"d"	"|d|"	"r"	"r±"
7.5	-0.5	0.5	1.0	-1.0
8.7	0.7	0.7	2.0	2.0
7.1	-0.9	0.9	3.0	-3.0
9.2	1.2	1.2	4.0	4.0
6.3	-1.7	1.7	5.0	-5.0
9.8	1.8	1.8	6.0	6.0
5.5	-2.5	2.5	7.5	-7.5
10.5	2.5	2.5	7.5	7.5
5.2	-2.8	2.8	9.0	-9.0
10.9	2.9	2.9	10.0	10.0
11.1	3.1	3.1	11.0	11.0
11.8	3.8	3.8	12.0	12.0
12.7	4.7	4.7	13.0	13.0
14.4	6.4	6.4	14.0	14.0

Here d_k denotes the difference $x_k - 8$. The realized statistic is

$$w = \sum_{k=1}^{14} r_k^{\pm} = 54$$

Under H_0, we have

$$T := \frac{W}{\sqrt{n(n+1)(2n+1)/6}} = \frac{W}{31.859} \overset{\text{approx}}{\sim} N(0,1)$$

by Lyapunov CLT. Without the whole-unit correction we calculate the realized value

$$t = 1.695$$

The approximate p-value, without correction, is therefore

$$P(Z \leq 1.695; H_0) > 0.5 > \alpha$$

Hence we fail to reject H_0.

iii.

Since the underlying distribution is symmetric, $m = \mu$. Our hypotheses can be rewritten as

$$\begin{aligned} H_0 &: \mu = 8 \\ H_1 &: \mu < 8 \end{aligned}$$

The test statistic is

$$T := \frac{\bar{X} - 8}{S/\sqrt{14}} \overset{\text{approx}}{\sim} t(13)$$

under H_0. The realized value is

$$t = 5.444$$

The p-value is given approximately by

$$P(t(13) \leq 5.444; H_0) > 0.5 > \alpha$$

We again fail to reject H_0.

2

(a)

Exclude the sample of exactly 9 blueberries.

Test statistic:

$$N^+ = \sum_{i=1}^{14} 1_{(X_i) > 9} \overset{H_0}{\sim} \text{Bin}(14, 0.5)$$

The realized value is

$$n^+ = 9$$

The p-value is therefore

$$P(N^+ \geq 9; H_0) = \sum_{n=9}^{14} \binom{14}{n} (0.5)^{14} = 0.21 > \alpha$$

We fail to reject H_0 at $\alpha = 0.1$.

(b)

Exclude the sample of exactly 9 blueberries.

Test statistic:

$$W = \sum_{k=1}^{14} R_k^{\pm}$$

where $R_k^{\pm} = \text{sign}(X_k - 9) \cdot R_k$ is the k-th signed rank.

"x"	"d"	"|d|"	"r"	"r±"
8.0	-1.0	1.0	3.0	-3.0
8.0	-1.0	1.0	3.0	-3.0
10.0	1.0	1.0	3.0	3.0
10.0	1.0	1.0	3.0	3.0
10.0	1.0	1.0	3.0	3.0
7.0	-2.0	2.0	7.0	-7.0
11.0	2.0	2.0	7.0	7.0
11.0	2.0	2.0	7.0	7.0
6.0	-3.0	3.0	10.5	-10.5
12.0	3.0	3.0	10.5	10.5
12.0	3.0	3.0	10.5	10.5
6.0	-3.0	3.0	10.5	-10.5
13.0	4.0	4.0	13.0	13.0
15.0	6.0	6.0	14.0	14.0

Here d_k denotes the difference $x_k - 9$. We have the realized statistic

$$w = 37$$

By CLT the p-value is given by (without the whole-unit correction)

$$\begin{aligned} P(W \geq 37; H_0) &= 1 - P(W \leq 35; H_0) \\ &\approx 1 - P\left(Z \leq \frac{35}{\sqrt{14(14+1)(2\cdot 14+1)/6}}\right) \\ &= 1 - P(Z \leq 1.0986) \\ &= 0.136 > \alpha \end{aligned}$$

We fail to reject H_0 at $\alpha = 0.1$.

(c)

Since the underlying distribution is symmetric, $m = \mu$. Our hypotheses can be rewritten as

$$\begin{aligned} H_0 &: \mu = 9 \\ H_1 &: \mu < 9 \end{aligned}$$

The test statistic is

$$T := \frac{\bar{X} - 9}{S/\sqrt{15}} \underset{\text{approx}}{\overset{H_0}{\sim}} t(14)$$

The realized value is

$$t = 4.660$$

The approximate p-value is given by

$$P(t(14) \geq 4.660; H_0) = 0.0002 < \alpha$$

We reject H_0 at 0.1.

3

We have the PDF and CDF for the uniform distributed X

$$\begin{aligned} f_X(x) &= 1/\gamma \\ F_X(x) &= x/\gamma \end{aligned}$$

Also by Lecture Note 12-1,

$$F_{X_{(k)}}(x) = \sum_{l=k}^{n} \binom{n}{l} F_X(x)^l (1 - F_X(x))^{n-l}$$

(a)

$$
\begin{aligned}
P\left(\frac{X_{(n)}}{\gamma} \le x\right) &= P\left(X_{(n)} \le \gamma x\right) \\
&= F_{X_{(n)}}(\gamma x) \\
&= F_X(\gamma x)^n \\
&= x^n
\end{aligned}
$$

(b)

$$
\begin{aligned}
P\left(X_{(n)} \le \gamma \le \frac{X_{(n)}}{\alpha^{1/n}}\right) &= P\left(\frac{X_{(n)}}{\gamma} \le 1 \le \frac{X_{(n)}}{\gamma}/\alpha^{1/n}\right) \\
&= P\left(\alpha^{1/n} \le \frac{X_{(n)}}{\gamma} \le 1\right) \\
&= P\left(\frac{X_{(n)}}{\gamma} \le 1\right) - P\left(\frac{X_{(n)}}{\gamma} \le \alpha^{1/n}\right) \\
&= 1^n - (\alpha^{1/n})^n \\
&= 1 - \alpha
\end{aligned}
$$

(c)

By CLT

$$
\frac{\bar{X} - \gamma/2}{\sigma_X/\sqrt{n}} \overset{\text{approx}}{\sim} N(0,1)
$$

And so

$$
\begin{aligned}
1 - \alpha &\approx P\left(-z_{\alpha/2} \le \frac{\bar{X} - \gamma/2}{\sigma_X/\sqrt{n}} \le z_{\alpha/2}\right) \\
&= P\left(2\bar{X} - \frac{2z_{\alpha/2}\sigma_X}{\sqrt{n}} \le \gamma \le 2\bar{X} + \frac{2z_{\alpha/2}\sigma_X}{\sqrt{n}}\right)
\end{aligned}
$$

However $\sigma_X = \gamma/\sqrt{12}$ is dependent of γ. We further approximate it by its unbiased estimator, S, leaving

$$
P\left(2\bar{X} - \frac{2z_{\alpha/2}S}{\sqrt{n}} \le \gamma \le 2\bar{X} + \frac{2z_{\alpha/2}S}{\sqrt{n}}\right) = 1 - \alpha
$$

as desired.

4

Let p_1, p_2, p_3, p_4 be the probability of homicides happening in the four seasons respectively. We test

$$
H_0 : p_1 = p_2 = p_3 = p_4 = \frac{1}{4}
$$

using the statistic

$$
Q_3 = \sum_{k=1}^{4} \frac{(Y_i - n/4)^2}{n/4} = \sum_{k=1}^{4} \frac{(Y_i - 340.25)^2}{340.25} \overset{H_0}{\sim} \chi^2(3)
$$

Season	Observed	Predicted
Spring (p_1)	340	340.25
Summer(p_2)	395	340.25
Autumn (p_3)	358	340.25
Winter (p_4)	268	340.25

We calculate the realized statistic

$$q_3 = 25.078 > 11.34 = \chi^2_{\alpha=0.01}(3)$$

Thus we reject H_0 at $\alpha = 0.01$.

5

To avoid confusion we rename the log-returns as $X_{1\ldots n-1=8717}$.

(a)

Test statistic:

$$N^+ := \sum_{i=1}^{8717} 1_{(X_i>0)} \overset{H_0}{\sim} \mathrm{Bin}(8717, 0.5) \approx N(4358.5, 2179.25)$$

is realized at

$$n^+ = 4391$$

The p-value is approximately

$$P(N^+ \geq 4391; H_0) \approx P\left(Z \geq \frac{4391 - 4358.5}{\sqrt{2179.25}}\right) = P(Z \geq 0.696) = 0.24 > \alpha$$

Hence we fail to reject $H_0 : m = 0$ at $\alpha = 5\%$.

(b)

After deleting 208 samples with $x_i = 0$, we are left with $\tilde{x}_{1\ldots\tilde{n}=8509}$ non-zero samples.

Test statistic:

$$T = \frac{\tilde{W}}{\sqrt{\frac{\tilde{n}(\tilde{n}+1)(2\tilde{n}+1)}{6}}} = \frac{\sum_{k=1}^{\tilde{n}} \tilde{R}_k^{\pm}}{453205.78} \underset{\text{approx}}{\overset{H_0}{\sim}} N(0,1)$$

where $\tilde{R}_k^{\pm} = \mathrm{sign}(\tilde{X}_k - 0) \cdot \tilde{R}_k$ is the k-th signed rank for the zero-filtered $\tilde{X}_k$.

```
  "x̃"          "d̃"          "|d̃|"          "r̃"       "r̃±"
 6.05202e-5    6.05202e-5   6.05202e-5       1.0         1.0
-7.96319e-5   -7.96319e-5   7.96319e-5       2.0        -2.0
-7.9694e-5    -7.9694e-5    7.9694e-5        3.0        -3.0
 0.000116689   0.000116689  0.000116689      4.0         4.0
 0.000116976   0.000116976  0.000116976      5.0         5.0
 0.000137343   0.000137343  0.000137343      6.0         6.0
-0.000142247  -0.000142247  0.000142247      7.0        -7.0
 0.000142816   0.000142816  0.000142816      8.0         8.0
-0.000144569  -0.000144569  0.000144569      9.0        -9.0
 0.000146057   0.000146057  0.000146057     10.0        10.0
-0.00015501   -0.00015501   0.00015501      11.0       -11.0
 0.000157337   0.000157337  0.000157337     12.0        12.0
 ⋮
-0.12574      -0.12574      0.12574       8498.0     -8498.0
 0.132929      0.132929     0.132929      8499.0      8499.0
 0.146328      0.146328     0.146328      8500.0      8500.0
-0.147453     -0.147453     0.147453      8501.0     -8501.0
-0.15631      -0.15631      0.15631       8502.0     -8502.0
-0.159453     -0.159453     0.159453      8503.0     -8503.0
 0.16525       0.16525      0.16525       8504.0      8504.0
-0.169577     -0.169577     0.169577      8505.0     -8505.0
 0.170626      0.170626     0.170626      8506.0      8506.0
 0.178692      0.178692     0.178692      8507.0      8507.0
-0.206445     -0.206445     0.206445      8508.0     -8508.0
-0.358332     -0.358332     0.358332      8509.0     -8509.0
```

Here $\tilde{d}_k$ denotes the difference $\tilde{x}_k - 0$. We calculate the approximate realized statistic

$$t \approx \frac{1964793}{453285.67} = 4.34$$

The p-value is approximately given by

$$P(Z \geq 4.33) < 0.05 = \alpha$$

So we reject H_0 at 5%.

(c)

The test statistic is

$$T := \frac{\bar{X} - 0}{S/\sqrt{8717}} \underset{\text{approx}}{\overset{H_0}{\sim}} t(8716) \approx N(0, 1)$$

The realized value is

$$t = 177.11$$

The approximate p-value is given by

$$P(Z \geq 177.11) < 0.05 = \alpha$$

We reject H_0 at 5%.

(d)

$$H_0 : R_i \overset{\text{i.i.d.}}{\sim} N(0, 0.02^2)$$

Categories	Probability (H_0)	Observed	Expected
< -0.001	0.480	3949	4184.16
$[-0.001, -0.0004)$	0.012	122	104.604
$[-0.0004, 0)$	0.008	47	69.736
$[0, 0.0004)$	0.008	277	69.736
$[0.0004, 0.001)$	0.012	117	104.604
≥ 0.001	0.480	4205	4184.16

The test statistic is

$$q_5 = \sum_{i=1}^{6} \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i} = 641.11$$

which, under H_0, is drawn from $\chi^2(5)$ since there are $k = 6$ categories. We compare it with

$$\chi^2_{0.05}(5) = 11.07 < q_5$$

Thus we reject H_0.

(e)

We have

$$\hat{\mu} = \bar{x} = 0.00089$$
$$\hat{\sigma^2} = s^2 = 0.00047$$

The null hypothesis:

$$H_0 : R_i \overset{\text{i.i.d.}}{\sim} N(\hat{\mu}, \hat{\sigma^2})$$

Categories	Probability (H_0)	Observed	Expected
< -0.001	0.4651	3949	4054.277
$[-0.001, -0.0004)$	0.011	122	95.887
$[-0.0004, 0)$	0.0074	47	64.506
$[0, 0.0004)$	0.0074	277	64.506
$[0.0004, 0.001)$	0.0111	117	96.759
≥ 0.001	0.4979	4205	4340.194

The test statistic is

$$q_7 = \sum_{i=1}^{6} \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i} = 723.04$$

which, under H_0, is drawn from $\chi^2(7)$ since there are $k = 6$ categories and $d = 2$ estimated parameters. We compare it with

$$\chi^2_{0.05}(7) = 14.07 < q_7$$

We reject H_0 at $\alpha = 0.05$.