A0235422N, Xiao Zhehao

1\. a) $F_Y(y)=\int_{-\infty}^{y} f_Y(y)\,dy=\int_0^y 2y\,dy=y^2,\ y\in(0,1)$ $\quad f_{Y_{(10)}}(y)=\frac{10!}{9!\,1!\,0!}[F_Y(y)]^9 f_Y(y)\,[1-F_Y(y)]^0=10(y^2)^9\,2y=20y^{19},\ y\in(0,1)$

b) $P(Y_{(10)}>0.9)=1-P(Y_{(10)}\le 0.9)=1-F_{Y_{(10)}}(0.9)=1-\binom{10}{10}[F_Y(0.9)]^{10}=1-[(0.9)^2]^{10}=0.878$

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

2\. a) $\hat{\pi}_p=X_{(a)}+b(X_{(a+1)}-X_{(a)})$, $a=\lfloor p(n+1)\rfloor$ & $b=p(n+1)-a$. $\quad n=14\quad a_{0.25}=\lfloor 0.25\cdot 15\rfloor=3\quad b_{0.25}=0.75\quad \hat{\pi}_{0.25}=X_3+0.75(X_4-X_3)=6.3+0.75(7.1-6.3)=6.9$

$a_{0.35}=5\quad b_{0.35}=0.25\quad \hat{\pi}_{0.35}=X_5+0.25(X_6-X_5)=7.5+0.25(8.7-7.5)=7.8\qquad a_{0.5}=7\quad b=0.5\quad \hat{m}=\hat{\pi}_{0.5}=X_7+0.5(X_8-X_7)=9.2+0.5(9.8-9.2)=9.5$

$a_{0.75}=11\quad b_{0.75}=0.25\quad \hat{\pi}_{0.75}=X_{11}+0.25(X_{12}-X_{11})=11.1+0.25(11.8-11.1)=11.275$

b) $P(X_{(i)}<\pi_p<X_{(j)})=\sum_{k=i}^{j-1}\binom{n}{k}p^k(1-p)^{n-k}=\sum_{k=0}^{j-1}\binom{n}{k}p^k(1-p)^{n-k}-\sum_{k=0}^{i-1}\binom{n}{k}p^k(1-p)^{n-k}=F_{Bin}(j-1;p)-F_{Bin}(i-1;p)$

i) $F_{Bin}(4;0.25)-F_{Bin}(0;0.25)=0.7415-0.0178=0.7237$ $\quad$ Hence, $(5.2, 7.5)$ is a 72.37% CI for $\pi_{0.25}$

ii) $F_{Bin}(7;0.35)-F_{Bin}(2;0.35)=0.9247-0.0839=0.8408$ $\quad$ Hence, $(6.3, 9.8)$ is a 84.08% CI for $\pi_{0.35}$

iii) $F_{Bin}(9;0.25)-F_{Bin}(4;0.25)=0.9102-0.0898=0.8204$ $\quad$ Hence, $(7.5, 10.9)$ is a 82.04% CI for $\pi_{0.5}$

c) i) $N^-=5$. Larger values of N^- favor H_1 $\quad P(N^-\ge 5;H_0)=P(Bin(14,0.5)\ge 5)=1-\overset{P(Bin(14,0.5)\le 4)}{0.0898}=0.9102>0.05=\alpha$ $\quad$ Hence, we dont reject H_0 at $\alpha=0.05$

ii)

X_i	5.2	11.8	7.1	14.4	6.3	7.5	11.1	5.5	9.8	12.7	10.5	10.9	8.7	9.2
X_i-m_0	-2.8	3.8	-0.9	6.4	-1.7	-0.5	3.1	-2.5	1.8	4.7	2.5	2.9	0.7	1.2
$\lvert X_i-m_0\rvert$	2.8	3.8	0.9	6.4	1.7	0.5	3.1	2.5	1.8	4.7	2.5	2.9	0.7	1.2
Ranks	9	12	3	14	5	1	11	7.5	6	13	7.5	10	2	4
Signed Ranks	-9	12	-3	14	-5	-1	11	-7.5	6	13	7.5	10	2	4

$W=54$ $\quad P(W\le 54;H_0)\approx\Phi\left(\frac{54+1}{\sqrt{\frac{14\cdot 15\cdot 29}{6}}}\right)=\Phi(1.726)\approx 0.9582>0.05=\alpha$

Assuming X_i are on a symmetric & continuous distribution, we dont reject H_0 at $\alpha=0.05$

iii) Assuming symmetric distribution, $\mu=m$ so we can test $H_0:\mu=8$, $H_1:\mu<8$. $\bar{x}=9.3357$ $\quad s^2=7.6917$ $\quad t=\frac{9.3357-8}{\sqrt{7.6917/14}}=1.802>1.771=t_{0.05}(13)$. Hence, we reject H_0 at $\alpha=0.05$

3\.

6	1.5	12.5	4.5	15	3	4.5	12.5	1.5	8	14	10.5	10.5	8	8
9	6	12	8	15	7	8	12	6	10	13	11	11	10	10

a) $N^+=9$ Larger values of N^+ favor H_1. $P(N^+\ge 9)=P(Bin(15,0.5)\ge 9)=1-P(Bin(15,0.5)\le 8)=1-0.6964=0.3036>0.1=\alpha$ Hence, we don't reject H_0 at $\alpha=0.1$

b)

X_i	~~9~~	6	12	8	15	7	8	12	6	10	13	11	11	10	10
X_i-m_0	~~0~~	-3	3	-1	6	-2	-1	3	-3	1	4	2	2	1	1
$\lvert X_i-m_0\rvert$	~~0~~	3	3	1	6	2	1	3	3	1	4	2	2	1	1
Ranks		10.5	10.5	3	14	7	3	10.5	10.5	3	13	7	7	3	3
Signed Ranks		-10.5	10.5	-3	14	-7	-3	10.5	-10.5	3	13	7	7	3	3

First sample removed as $\lvert X_i-m_0\rvert=0$. $W=37$

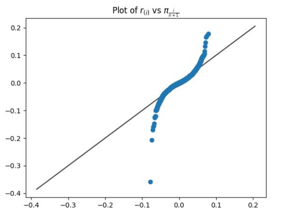
$P(W\ge 37;H_0)=1-P(W\le 35;H_0)\approx 1-\Phi\left(\frac{35+1}{\sqrt{\frac{14\cdot 15\cdot 29}{6}}}\right)=1-\Phi(1.099)\approx 1-0.8643=0.1357>0.1=\alpha$

Assuming X_i are on a symmetric & continuous distribution, we don't reject H_0 at $\alpha=0.1$

c) Assuming symmetric distribution, $\mu=m$ so we can test $H_0:\mu=9$, $H_1:\mu>9$. $\bar{x}=9.8667$ $\quad s^2=6.6952$ $\quad t=\frac{9.8667-8}{\sqrt{6.6952/15}}=1.297<1.345=t_{0.1}(14)$. Hence, we don't reject H_0 at $\alpha=0.1$

4\. a) $\bar{r}=0.000839$ $\quad s_r^2=0.000445$

b)



c) r_i does not follow an $N(\bar{r},s_r^2)$ distribution as the qqplot shows $\bar{r}$ is greater than $\pi_{0.5}$ and the tails are lighter than expected.

5\. a) $N^+=4960$ Larger values of N^+ favor H_1. $P(N^+\ge 4960)=P(Bin(15,0.5)\ge 4960)=0.164>0.05=\alpha$ Hence, we don't reject H_0 at $\alpha=0.05$

b) Assuming symmetric distribution, $\mu=m$ so we can test $H_0:\mu=0$, $H_1:\mu>0$. Since n is large, we can approximate t with z $\quad z=\frac{0.000839}{\sqrt{0.000445/9822}}=3.939>1.645=z_{0.05}$

Hence, we reject H_0 at $\alpha=0.05$

c) $W=2555960$ $\quad P(W\ge 2555960)=2.713\times 10^{-6}<0.05=\alpha$ $\quad$ Assuming X_i are on a symmetric & continuous distribution, we reject H_0 at $\alpha=0.05$

ST2132 Mathematical Statistics
Assignment 4
Due date: 28 March 2025 23:59 on Canvas
Please submit in pdf format.

Please work on the assignment by yourself only. We follow a strict rule on academic honesty.

Generative AI policy: The use of generative AI tools (e.g., ChatGPT, Gemini, etc.) is strictly prohibited for structural or conceptual questions in this course. However, these tools may be used to assist with coding tasks, provided their use is clearly documented.

1. The proportion of rats that successfully complete a designed experiment (e.g., running through a maze) is of interest for psychologists. Denote by Y the proportion of rats that complete the experiment, and suppose that the experiment is replicated in 10 different rooms. Assume that $Y_1, Y_2, \ldots, Y_{10}$ are i.i.d. Beta random variables with $\alpha = 2$ and $\beta = 1$. Recall that for this Beta model, the pdf is

$$f_Y(y) = \begin{cases} 2y & \text{if } 0 < y < 1\,, \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Find the pdf of $Y_{(10)}$, the largest order statistic.

   (b) Calculate $P(Y_{(10)} > 0.9)$.

2. Let X be the weights of newborn babies in pounds, and denote m to be its median. For a sample of 14, we have:

$$5.2 \quad 11.8 \quad 7.1 \quad 14.4 \quad 6.3 \quad 7.5 \quad 11.1 \quad 5.5 \quad 9.8 \quad 12.7 \quad 10.5 \quad 10.9 \quad 8.7 \quad 9.2$$

   (a) Give point estimates of $\pi_{0.25}, \pi_{0.35}, m, \pi_{0.75}$.

   (b) Find the following confidence intervals and, from Table II in Appendix B, state the associated confidence coefficient:

      i. $(X_{(1)}, X_{(5)})$, a confidence interval for $\pi_{0.25}$.

      ii. $(X_{(3)}, X_{(8)})$, a confidence interval for $\pi_{0.35}$.

      iii. $(X_{(5)}, X_{(10)})$, a confidence interval for $\pi_{0.5}$.

   (c) Now, we would like to test $H_0 : m = 8$ against $H_1 : m < 8$ at significance level $\alpha = 0.05$.

      i. Use the sign test to test the hypothesis.

      ii. Use the Wilcoxon sign-rank test to test the hypothesis.

      iii. Use the t-test to test the hypothesis, assuming the underlying distribution is symmetric.

3. Let X be the number of blueberries on a muffin made by Michael, and denote by m to be its median. Charles suspects that Michael has been adding more blueberries than usual to his muffins. For some reason, he has kept meticulous records. The last 15 muffins Charles bought contained the following numbers of blueberries:

$$9 \quad 6 \quad 12 \quad 8 \quad 15 \quad 7 \quad 8 \quad 12 \quad 6 \quad 10 \quad 13 \quad 11 \quad 11 \quad 10 \quad 10$$

   Charles would like to test $H_0 : m = 9$ against $H_1 : m > 9$ at significance level $\alpha = 0.1$.

   (a) Use the sign test to test the hypothesis.

(b) Use the Wilcoxon sign-rank test to test the hypothesis. (Note: Assume that the distribution is continuous)

(c) Use the t-test to test the hypothesis, assuming the underlying distribution is symmetric.

4. (Does Microsoft stock price really follow a lognormal distribution?) In finance literature, one popular model for stock price is the lognormal distribution. Suppose that we have the daily stock price of a stock, say $S_1, S_2, \ldots, S_n$. We can then compute the so-called log-return of the stock price $R_1, R_2, \ldots, R_{n-1}$, where

$$R_i = \ln\left(\frac{S_{i+1}}{S_i}\right), \quad i = 1, \ldots, n-1.$$

On Canvas, there is a dataset "MSFT.csv" that contains the daily stock price of Microsoft from March 14 1986 to March 10 2025, with $n = 9823$. Use your favourite computing language to answer this question, and attach your code at the end of your assignment.

(a) Compute r_i of the dataset, the realizations of R_i for $i = 1, \ldots, 9822$. Report the sample mean $\overline{r}$ and the sample variance s_r^2.

(b) Draw a QQ plot of the sample quantiles of r_i against the theoretical quantiles of $N(\overline{r}, s_r^2)$. Label the axis on your plot clearly. **Do not use any existing packages or functions in your computing language that automatically draw the QQ plot for you. Instead, you can use any other functions or packages to determine the sample quantiles and the theoretical quantiles, i.e. try creating the QQ plot from scratch using only sample quantiles and theoretical quantiles.**

(c) Comment on the QQ plot. Do r_i really follow a normal distribution based on your plot?

5. (Microsoft stock price continued) **Do not use any existing packages or functions in your computing language that automatically conduct the tests for you. Instead, you can use any other functions or packages to help you conduct these tests.**

(a) From the QQ plot, it seems that the log-return r_i is symmetric about the origin. Test $H_0 : m = 0$ against $H_1 : m > 0$ using the sign test at $\alpha = 5\%$.

(b) Use the t-test to test the hypothesis in part (a).

(c) (**Optional, for fun!**) Use the Wilcoxon sign-rank test to test the hypothesis in part (a). This is data with ties and with data such that $r_i = 0$, so remember to clean the data! In Excel, use the function RANK.AVG to compute the average rank. In Python, you can import Pandas and use dataframe.rank() to compute the average rank. In R, you can use the function rank() to compute the average rank. In MATLAB, you can use the function tiedrank() to compute the average rank.