Problem 2:

Sort Method A

1 2 3 4 5 6 7 8 9 10 11 12 13

79.97, 79.98, 80.00, 80.02, 80.02, 80.02, 80.03, 80.03, 80.03, 80.04, 80.04, 80.04, 80.05

$Q1 = \text{Method A sorted}\left[\frac{13+1}{4}\right] = \frac{80.00+80.02}{2} = 80.01$

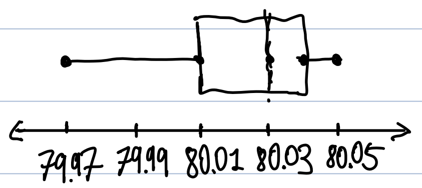
$Q3 = \text{Method A sorted}\left[\frac{3(13+1)}{4}\right] = \frac{80.04+80.04}{2} = 80.04$

$IQR = 0.03$

$\text{Median} = \text{Method A sorted}\left[\frac{13+1}{2}\right] = 80.3$

Min = 79.97

Max = 80.05



Sort Data

1 2 3 4 5 6 7

Method B sorted = 79.94, 79.95, 79.97, 79.97, 79.98, 80.02, 80.03

$Q_1 = \text{Method B sorted}\left[\frac{7+1}{4}\right] = 79.95$

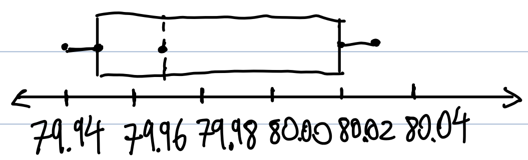
$Q_3 = \text{Method B sorted}\left[\frac{3(7+1)}{4}\right] = 80.02$

$IQR = 0.03$

$\text{Median} = \text{Method A sorted}\left[\frac{7+1}{2}\right] = 79.97$

Min = 79.94

Max = 80.03



Histogram Method A

79.97 – 79.99 : 2

80.00 – 80.02 : 4

80.03 – 80.05 : 7

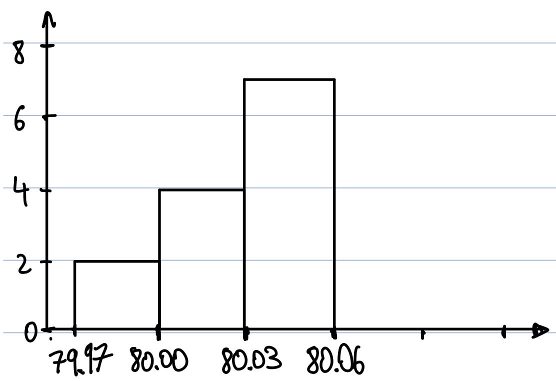
Histogram Method B

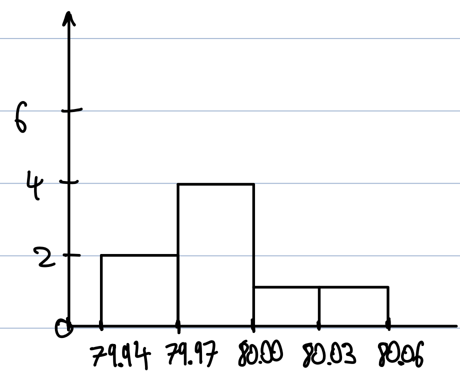
79.4 – 79.6 : 2

79.7 – 79.9 : 4

80.0 – 80.02 : 1

80.03 – 80.05 : 1





Method A QQ Plot

$z = \frac{x-\mu}{\sigma}$ where $\mu_A = 80.02077$, $\mu_B = 79.97875$

$\sigma_A = 0.02396579$ $\sigma_B = 0.03136764$

Z values for Method A Sorted

$(79.97 - 80.02)/0.024 = -2.086$

$(79.98 - 80.02)/0.024 = -1.670$

$(80.00 - 80.02)/0.024 = -0.835$

$(80.02 - 80.02)/0.024 = 0.000$

$(80.02 - 80.02)/0.024 = 0.000$

$(80.02 - 80.02)/0.024 = 0.000$

$(80.03 - 80.02)/0.024 = 0.4173$

$(80.03 - 80.02)/0.024 = 0.4173$

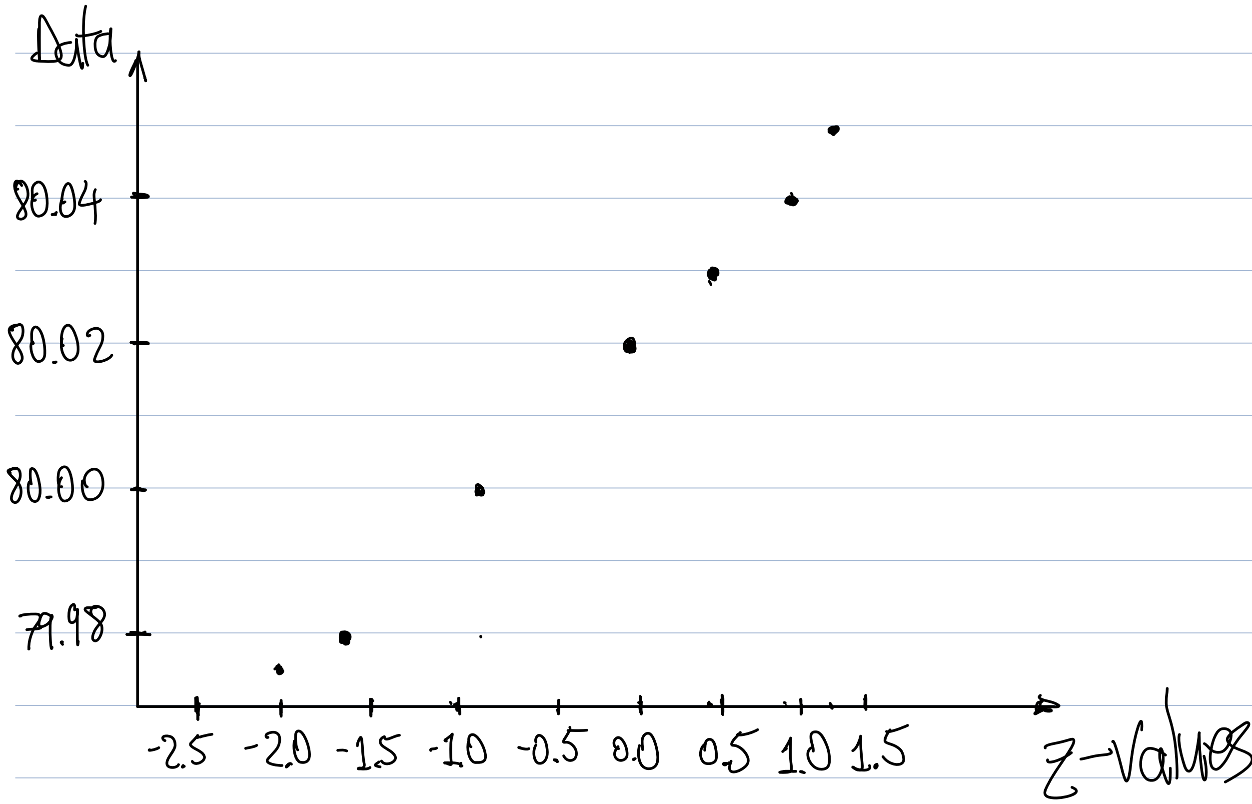
$(80.03 - 80.02)/0.024 = 0.4173$

$(80.04 - 80.02)/0.024 = 0.8345$

$(80.04 - 80.02)/0.024 = 0.8345$

$(80.04 - 80.02)/0.024 = 0.8345$

$(80.05 - 80.02)/0.024 = 1.252$

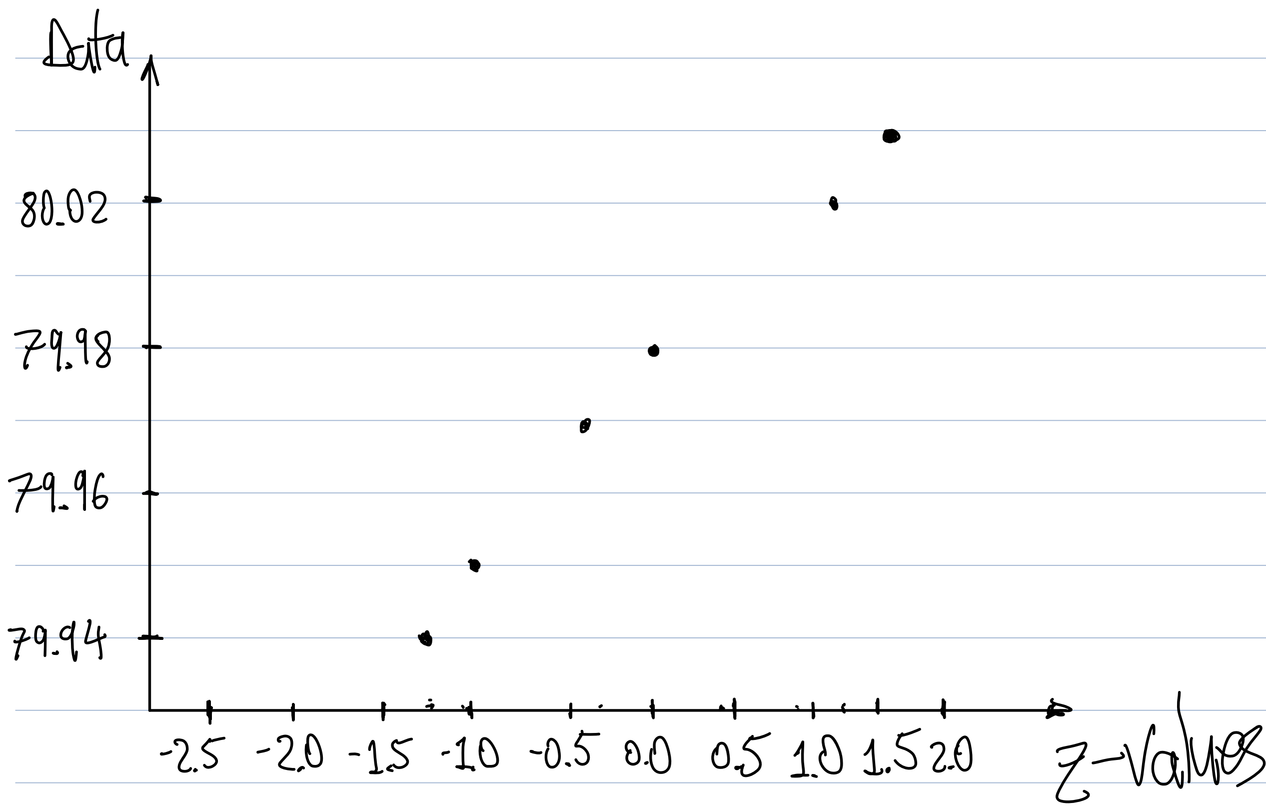


Method B QQPlot

$\sigma_A = 0.0314$ $\mu_B = 79.98$

Method B sorted = 79.94, 79.95, 79.97, 79.97, 79.98, 80.02, 80.03
(labelled 1 to 7 above the values)

$(79.94 - 79.98)/0.0314 = -1.2752$
$(79.95 - 79.98)/0.0314 = -0.95164$
$(79.97 - 79.98)/0.0314 = -0.3188$
$(79.97 - 79.98)/0.0314 = -0.3188$
$(79.97 - 79.98)/0.0314 = -0.3188$
$(79.98 - 79.98)/0.0314 = 0.000$
$(80.02 - 79.98)/0.0314 = 1.2752$
$(80.03 - 79.98)/0.0314 = 1.594$



Problem 2b equations:

(1) Equal Mean, Equal Variance → Studentized t-test

$$\frac{\bar{X}_1 - \bar{X}_2}{s_p\left(\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)}$$

$\bar{X}_1, \bar{X}_2$ = two sample means

n_1, n_2 = sample sizes for sample 1 and sample 2.

$$s_p = \sqrt{(n_1-1)s_1^2+(n_2-1)s_2^2/(n_1+n_2-2)}$$

2). Equal Mean, Unequal Variance → Welch's t-test

$$\frac{(\bar{X}_1 - \bar{X}_2)}{\left(\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)}$$

$\bar{X}_1, \bar{X}_2$ are sample means

n_1, n_2 are the samples sizes

c). Wilcoxon/Mann-Whitney Test

Test stat is $\min(U_1, U_2)$ compared aginst threshold

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where R_1, R_2 are the respective sum of Ranks for each group

[illegible]

Problem 3

Median

$-1.43, -0.95, -0.19, 0.02, 0.14, 0.83, 1.35, 1.46, 2.62$

1 2 3 4 5 6 7 8 9

(a). Calculate the sample median, IQR (Interquartile Range) and MAD (Median Absolute Deviation); What are the (robust) breakdown points of these statistics?

• Median: The numbers are sorted so we have the median as the $\frac{n+1}{2}$ term which is $\boxed{0.14}$ where $n=9$.

• IQR:

IQR = Q3 − Q1 where Q1 is the median of first half, and Q3 is median of second half.

$$Q1 = \frac{-0.95+(-0.19)}{2} = -0.57 \qquad Q_3 = \frac{1.35+1.46}{2} = 1.405$$

$$IQR = 1.405-(-0.57) = \boxed{1.975}$$

• MAD

$$MAD = \text{median}(|X_i - \tilde{X}|) \quad \text{where } \tilde{X} = \text{median}(X)$$

X shifted by Median

$-1.57, -1.09, -0.33, -0.12, 0.00, 0.69, 1.21, 1.32, 2.48$

Absolute Value

1.57, 1.09, 0.33, 0.12, 0.00, 0.69, 1.21, 1.32, 2.48

Sorted

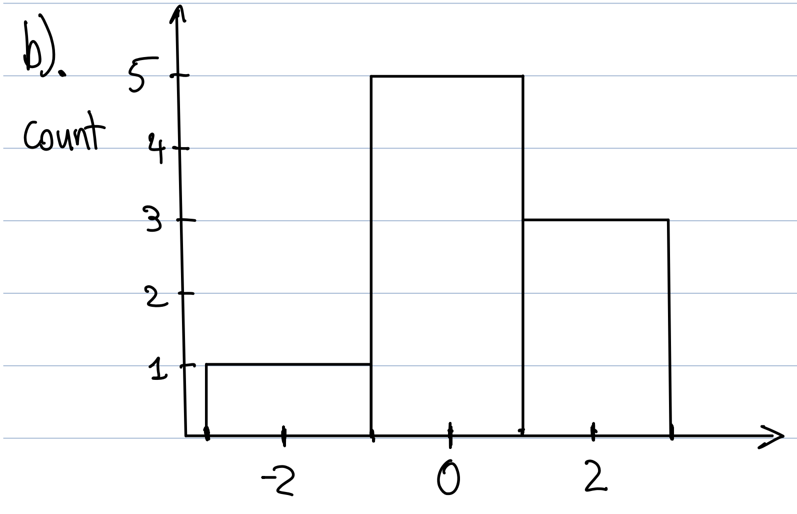
0.00, 0.12, 0.33, 0.69, 1.09, 1.21, 1.32, 1.57, 2.48

Thus MAD = 1.09

Robust Breakdown Points:

- Median: 50% because up to 50% of Data Points can be changed without changing the Median
- IQR: 25% because if we change just 1/4 of the data, say in the bottom quarter, Q1 will change and so will the IQR
- MAD: Since MAD is a function of the Median, changing 50% of the values will change the value that the data is shifted by.

b).



Problem 4:

Consider the model $y_i = B_0 + x_i B_i + \epsilon_i$ where ϵ_i i.i.d $N(0, \sigma^2)$ for $i = 1, 2, \ldots, 5$. We have the following data:

X	Y
-1.0	-1.5
0.0	0.3
1.0	0.9
2.0	2.1
3.0	2.3

a). Find the Least Squares (LS) and MLE estimates of B_1, σ^2 and variance of $\hat{B}_1$.

$$\bar{X} = \frac{-1.0+0.0+1.0+2.0+3.0}{5} = 1.0$$

$$\hat{B}_1 = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}$$

$$\bar{Y} = \frac{-1.5+0.3+0.9+2.1+2.3}{5} = 0.82$$

$$\hat{B}_1 = \frac{(-1.0-1.0)(-1.5-0.82)+(0.0-1.0)(0.3-0.82)+(1.0-1.0)(0.9-0.82)+(2.0-1.0)(2.1-0.82)+(3.0-1.0)(2.3-0.82)}{(-1.0-1.0)^2+(0.0-1.0)^2+(1.0-1.0)^2+(2.0-1.0)^2+(3.0-1.0)^2}$$

$$= \frac{(-2)(-2.32)+(-1)(-0.52)+0+(1.0)(1.28)+(2.0)(1.48)}{4.0+1.0+0+1+4} =$$

$$= \frac{4.64+0.52+1.28+2.96}{10} = 0.94$$

$$\hat{B}_0 = \bar{Y} - B_1\bar{X} = 0.82 - 0.94(1.0) = -0.12$$

$$\sigma^2 \approx s^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} = \frac{(2.32)^2 + (0.52)^2 + (0.08)^2 + (1.28)^2 + (1.48)^2}{5-1} = \frac{9.488}{4} = 2.372$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2} = \frac{2.372}{10} = 0.2372$$

<u>MLE estimate</u>

We are working with a Normal Dist where

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

Our likelihood function is:

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right\}$$

The log likelihood is:

$$\log L(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 + \text{const}$$

Take first derivative and set right hand side to 0

$$(\hat{\beta}_0^{MLE}, \hat{\beta}_1^{MLE}) = \underset{\beta_0\, \beta_1}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\begin{cases} \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \\ \beta_0 n + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \end{cases}$$

Solve the System of Equations:

$$\hat{\beta}_1^{MLE} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\hat{\beta}_0^{MLE} = \bar{y} - \hat{\beta}_1^{MLE}\bar{x}$$

We note that the equations are the exact same as for the least square estimates

Thus $\hat{\beta}_1^{MLE} = \hat{\beta}_1^{LS} = 0.94$ and $\hat{\beta}_0^{MLE} = \hat{\beta}_0^{LS} = -0.12$

$$\sigma^2_{MLE} = \frac{\sum_{i=1}^{n}\{y_i - (\hat{\beta}_0^{MLE} + \hat{\beta}_1^{MLE} x_i)\}^2}{n}$$

$$= \frac{\sum_{i=1}^{n}\{y_i - (-0.12 + 0.94 x_i)\}^2}{5}$$

$$= \frac{(-1.5-(-0.12+0.94(-1.0)))^2}{5} + \frac{(0.3-(-0.12+(0)))^2}{5} +$$

$$\frac{(0.9-(-0.12+(1.0)(0.94)))^2}{5} + \frac{(2.1-(-0.12+2.0(0.94)))^2}{5} + \frac{(2.3-(-0.12+3.0(0.94)))^2}{5}$$

$$= 0.03872 + 0.03528 + 0.00128 + 0.2312 + 0.032 = \boxed{0.134}$$

$$\mathrm{Var}(\hat{\beta}_1^{MLE}) = \frac{\sigma}{\sum (x_i - \bar{x})^2} = \frac{0.134}{10} = \boxed{0.0134}$$

b). Find the 95% confidence interval for β_1.

$\hat{\beta}_1 \pm t(1-0.05/2; 5-2) SE(\hat{\beta}_1)$

$SE(\hat{\beta}_1) = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{2.372}{5}} = 0.6888$

$0.94 \pm 3.18 * 0.6888 = [-1.2504, 3.1304]$

Problem 5: Suppose $X_1, \ldots, X_n \overset{iid}{\sim} Uniform[0, \theta]$. We would like to estimate and make inference for θ.

(a). Find the MLE $\hat{\theta}^{MLE}$ for θ and compute the density of $\hat{\theta}^{MLE}$.

For a uniform distribution, the likelihood function is:

$$\prod_{i=1}^{n} f(x_i; a, b) = \prod_{i=1}^{n} \frac{1}{(b-a)} = \frac{1}{(b-a)^n}$$

We then take the log to get the log-likelihood function.

$$\log \prod_{i=1}^{n} f(x_i; a, b) = \log \prod_{i=1}^{n} \frac{1}{(b-a)} = \log\left(\frac{1}{(b-a)^n}\right) = -n\log(b-a)$$

we want to find the derivative of log-likelihood function with respect to a and b.

w.r.t a: $\frac{n}{(b-a)}$

w.r.t b: $\frac{-n}{(b-a)}$

The derivative w.r.t a is monotonically increasing.

- The MLE estimate for a is $\min(X_1,\dots,X_n)=0$.
- The MLE estimate for b is $\max(X_1,\dots,X_n)$

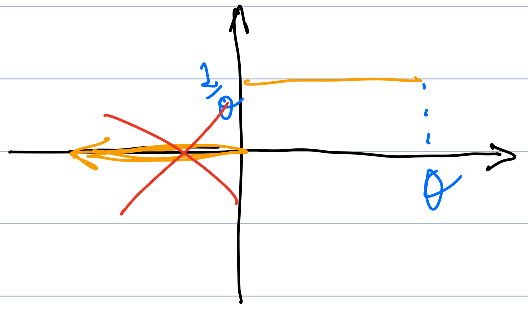
In our case $a=0$ and $b=\theta$.

Thus, the estimate for θ is $\boxed{\max(X_1,\dots,X_n)}$

The density function

$$f(x|\theta)=\begin{cases}\frac{1}{\theta}, & 0\le x<\theta\\ 0, & \text{otherwise}\end{cases}$$

(b). Let the Underlying truth be $\theta=1$. Suppose we are only able to get $n=60$



(c)

Based on Part (b) $P(\hat{\theta}^{MLE}=1)=0$.
1 is the asymptotic bound. and we never actually reach it.

To compute $P(\hat{\theta}^{*}=\hat{\theta}^{MLE})$