Normal Approximation to the Binomial Distribution

- In a digital communication channel, assume that the number of bits received in error can be modeled by a binomial random variable, and assume that the probability that a bit is received in error is 0.00001. If 16 million bits are transmitted, what is the probability that more than 150 errors occur?

$$P(X > 150) = 1 - P(x \le 150) = 1 - \sum_{x=0}^{150} \binom{16000000}{x} (10^{-5})^x (0.99999)^{16,000,000-x}$$

Normal Approximation to the Binomial Distribution II

- If X is a binomial random variable,

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

- Then, it is approximately a standard normal variable. The approximation is good for

$$np > 5 \quad \text{and} \quad n(1-p) > 5$$

- Continuity correction

$$P(X \le x) = P(X \le x + 0.5) \cong P\left(Z \le \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

$$P(x \le X) = P(x - 0.5 \le X) \cong P\left(Z \le \frac{x - 0.5 - np}{\sqrt{np(1-p)}} \le Z\right)$$

Example 4-18: Applying the Approximation

In a digital communication channel, assume that the number of bits received in error can be modeled by a binomial random variable. The probability that a bit is received in error is 10^{-5}. If 16 million bits are transmitted, what is the probability that 150 or fewer errors occur?

$$P(X \le 150) = P(X \le 150.5)$$

$$= P\left(\frac{X - 160}{\sqrt{160(1-10^{-5})}} \le \frac{150.5 - 160}{\sqrt{160(1-10^{-5})}}\right)$$

$$= P\left(Z \le \frac{-9.5}{12.6491}\right) = P(Z \le -0.75104) = 0.2263$$

1. 68% will fail between ±1σ

19 ± 1(1.2) = [17.8,20.2]

2. 95% will fail between ±2σ

19 ± 2(1.2) = [16.6,21.4]

3. Virtually all will fail at about ±3σ

19 ± 3(1.2) = [15.4,22.6]

Standard Normal Distribution

A normal random variable with

$$\mu = 0 \text{ and } \sigma^2 = 1$$

is called a standard normal random variable and is denoted as Z. The cumulative distribution function of a standard normal random variable is denoted as:

$$\Phi(z) = P(Z \le z)$$

Values are found in Appendix Table III and by using Excel and Minitab.

Excel functions available
- NORM.S.DIST(z) : standard normal distribution
- NORM.S.INV(p) : inverse standard normal distribution
- NORM.INV(p,m,s): inverse general
- NORM.DIST(X,m,s,Cumulative) : general

Standardizing a Normal Random Variable

- Suppose X is a normal random variable with mean μ and variance σ^2. Then,

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P(Z \le z) = \Phi(z)$$

- where Z is a **standard normal random variable,** and $Z = \frac{x-\mu}{\sigma}$ is the ***z*-value** obtained by **standardizing** X.
- The standard normal distribution has a mean of zero and a variance of one: $E(Z) = 0$ and $V(Z) = 1$.

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

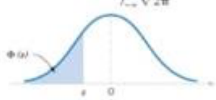


Table III Cumulative Standard Normal Distribution

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	−0.00
−3.9	0.000033	0.000034	0.000036	0.000037	0.000039	0.000041	0.000042	0.000044	0.000046	0.000048
−3.8	0.000050	0.000052	0.000054	0.000057	0.000059	0.000062	0.000064	0.000067	0.000069	0.000072
−3.7	0.000075	0.000078	0.000082	0.000085	0.000088	0.000092	0.000096	0.000100	0.000104	0.000108
−3.6	0.000112	0.000117	0.000121	0.000126	0.000131	0.000136	0.000142	0.000147	0.000153	0.000159
−3.5	0.000165	0.000172	0.000179	0.000185	0.000193	0.000200	0.000208	0.000216	0.000224	0.000233
−3.4	0.000242	0.000251	0.000260	0.000270	0.000280	0.000291	0.000302	0.000313	0.000325	0.000337
−3.3	0.000350	0.000362	0.000376	0.000390	0.000404	0.000419	0.000434	0.000450	0.000467	0.000483
−3.2	0.000501	0.000519	0.000538	0.000557	0.000577	0.000598	0.000619	0.000641	0.000664	0.000687
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

NOTE : The column headings refer to the hundredths digit of the value of z in $P(Z \le z)$. For example, $P(Z \le 1.53)$ is found by reading down the z column to the row 1.5 and then selecting the probability from the column labeled 0.03 to be 0.93699.

EXAMPLE 4-12

The following calculations are shown pictorially in Fig. 4-14. In practice, a probability is often rounded to one or two significant digits.

(1) $P(Z > 1.26) = 1 - P(Z \le 1.26) = 1 - 0.89616$
$= 0.10384$

(2) $P(Z < -0.86) = 0.19490$.

(3) $P(Z > -1.37) = P(Z < 1.37) = 0.91465$

(4) $P(-1.25 < Z < 0.37)$. This probability can be found from the difference of two areas, $P(Z < 0.37) - P(Z < -1.25)$. Now,

$P(Z < 0.37) = 0.64431$

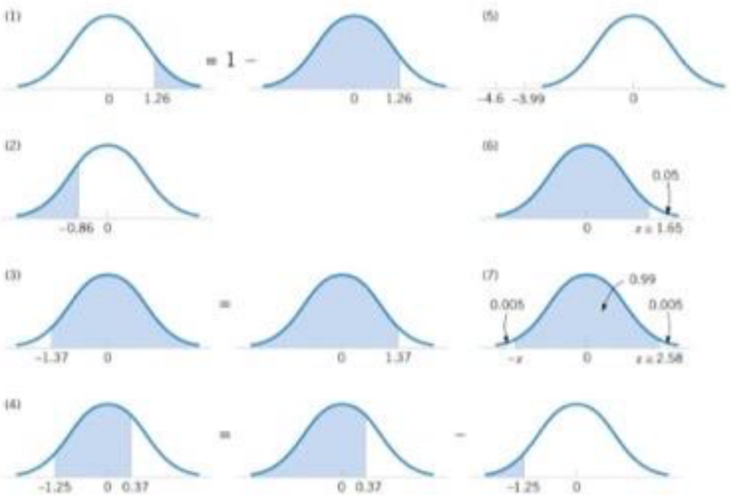
and

$P(Z < -1.25) = 0.10565$

Therefore,

$P(-1.25 < Z < 0.37) = 0.64431 - 0.10565$
$= 0.53866$

(5) $P(Z \le -4.6)$ cannot be found exactly from Appendix Table III. However, the last entry in the table can be used to find that $P(Z \le -3.99) = 0.00003$. Because $P(Z \le -4.6) < P(Z \le -3.99)$, $P(Z \le -4.6)$ is nearly zero.



Example 4-14: Normally Distributed Current-1

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with μ = 10 and σ = 2 mA, what is the probability that the current measurement is between 9 and 11 mA?

$$P(9 < X < 11) = P\left(\frac{9-10}{2} < \frac{x-10}{2} < \frac{11-10}{2}\right)$$
$$= P(-0.5 < z < 0.5)$$
$$= P(z < 0.5) - P(z < -0.5)$$
$$= 0.69146 - 0.30854 = 0.38292$$

−0.6	0.245097	0.248252	0.251429	0.254627	0.257846	0.261086	0.264347	0.267629	0.270931	[illegible]
−0.5	0.277595	0.280957	0.284339	0.287740	0.291160	0.294599	0.298056	0.301532	0.305026	0.308538
−0.4	0.312067	0.315614	0.319178	0.322758	0.326355	0.329969	0.333598	0.337243	0.340903	[illegible]
0.4	[illegible]	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903

Excel
= NORM.DIST(11,10,2,TRUE)-NORM.DIST(9,10,2,TRUE) =**0.382925**

Cashews & Nuts Example (Continuous)

A nut company markets cans of deluxe mixed nuts containing almonds, cashews, and peanuts. The net weight of the can is 1 lb, but the weight contribution of each type of nut is random. Because the three weights sum to 1, a joint probability model for any two gives all necessary information about the weight of the third type. Let X= the weight of almonds in a selected can and Y= weight of cashews. The joint PDF for (X,Y) was suggested to be

$$f(x,y)=\begin{cases}24xy & 0\le x\le 1, 0\le y\le 1, x+y\le 1\\ 0 & \text{otherwise}\end{cases}$$

1. Show that this function is indeed a PDF.
2. Are X and Y independent?
3. Find the marginal PDF of X and Y.
4. If the cost per lb of almonds is \$1.00, cashews \$1.5, and of peanuts is \$0.5. Find the expected cost of a one lb can of mixed nuts.
5. Find the covariance and correlation.

Example 2-8: Sampling w/o Replacement-1

- A bin of 50 parts contains 3 defectives and 47 non-defective parts. A sample of 6 parts is selected from the 50 without replacement. How many samples of size 6 contain 2 defective parts? What is the probability of having a sample of six parts with two defective parts.
- First, how many ways are there for selecting 2 parts from the 3 defective parts?

$$C_2^3=\frac{3!}{2!\cdot 1!}=3 \text{ different ways}$$

- Now, how many ways are there for selecting 4 parts from the 47 non-defective parts?

$$C_4^{47}=\frac{47!}{4!\cdot 43!}=\frac{47\cdot 46\cdot 45\cdot 44\cdot 43!}{4\cdot 3\cdot 2\cdot 1\cdot 43!}=178{,}365 \text{ different ways}$$

Example 2-8: Sampling w/o Replacement-2

- Now, how many ways are there to obtain:
  - 2 from 3 defectives, and
  - 4 from 47 non-defectives?

$$C_2^3 C_4^{47}=3\times 178{,}365=535{,}095 \text{ different ways}$$

- The total number of different subsets of size 6 is

$$C_6^{50}=\binom{50}{6}=\frac{50!}{6!44!}=15{,}890{,}700$$

- The probability of having six parts with two defective

$$P(\text{Two defective parts out of six})=\frac{C_2^3 C_4^{47}}{C_6^{50}}=\frac{535{,}095}{15{,}890{,}700}$$

In MATLAB
nchoosek(3,2) =3
nchoosek(47,4) = 178365

In EXCEL
COMBIN(3,2) =3
COMBIN(47,4) = 178365

Example on Bayes' Theorem

Known: Vendors I, II, III, and IV provide all the bushings to a certain factory. The following table shows the amounts supplied and percentage of good parts supplied by the vendors.
Wanted: What is the probability that a randomly selected bushing is bad? What is the probability it came from vendor 3?

Vendor	% Supply	% Good
I	25	80
II	35	95
III	10	70
IV	30	90

Solution

$$P(A)=\sum_{i=1}^{4}P(A\cap B_i)=\sum_{i=1}^{4}P(B_i)P(A\mid B_i)$$

$$P(A)=0.25(0.2)+0.35(0.05)+0.1(0.3)+0.3(0.1)$$

$$P(A)=0.1275$$

Therefore

$$P(B_3\mid A)=\frac{P(B_3\cap A)}{P(A)}=\frac{0.03}{0.1275}=0.2353$$

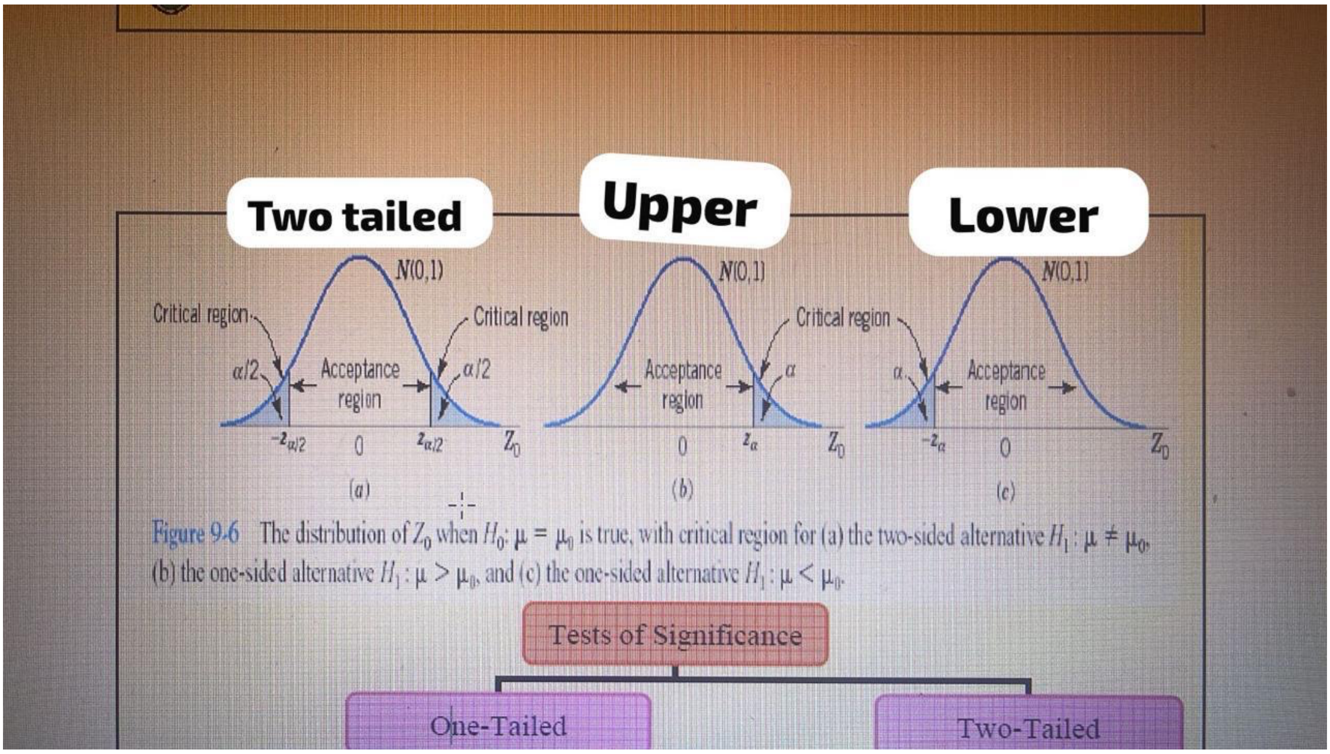
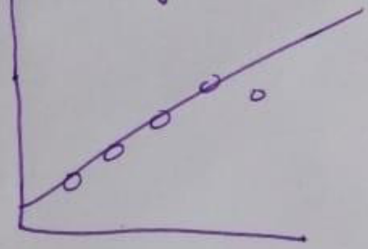


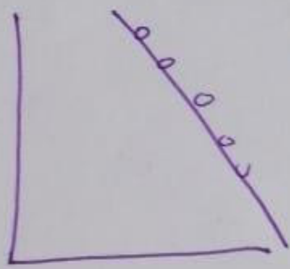
Figure 9-6 The distribution of Z_0 when H_0: $\mu=\mu_0$ is true, with critical region for (a) the two-sided alternative $H_1: \mu \neq \mu_0$, (b) the one-sided alternative $H_1: \mu > \mu_0$, and (c) the one-sided alternative $H_1: \mu < \mu_0$.

*

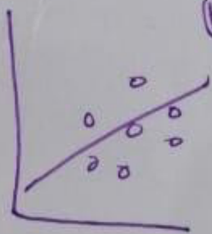
$R \simeq 1$



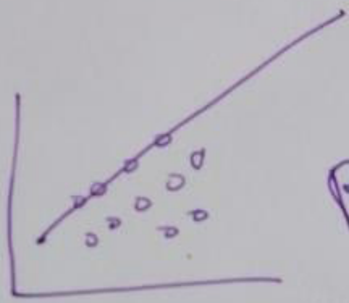
$R \simeq -1$

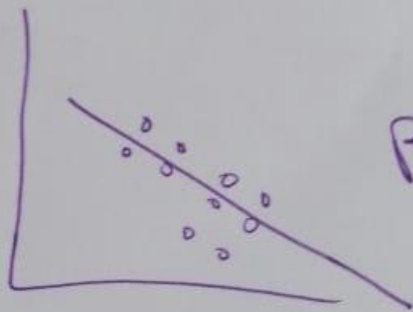


$R \simeq 0$



$R \simeq (0-1)$





$R \simeq (-1-0)$

*

1. (15 Marks) The Environmental Protection Agency (EPA) perform extensive tests on all new car models to determine their mileage ratings. The 25 measurements given below represent the results of the test on a sample of size 25 of a new car model.

31.8	33.1	33.9	33.9	34.8
35.5	36.3	36.5	36.7	36.9
37.0	37.0	13 37.1	37.1	37.3
37.6	38.5	38.6	39.0	40.2
40.3	40.5	41.0	41.0	44.9

These data has a mean of 37.460 with a variance of 8.382. Please answer/carry out the following: s^2

a) (1 mark) What is the median of these data? 37.1

b) (1 mark) What is the coefficient of variation of these data (%)? $\frac{S}{\bar{x}} \times 100 = \frac{\sqrt{8.382}}{37.46} = 0.077 = 7.7\%$

c) (1 mark) What is Q1 of these data? $(35.5 + 36.3)/2 = 35.9$

d) (1 mark) What is Q3 of these data? $(39 + 40.2)/2 = 39.6$

e) (1 mark) What is IQR of these data? $Q_3 - Q_1 = 3.7$

f) (1 mark) What is the skewness of these data? $\frac{3(\text{mean} - \text{median})}{\text{Stand. deviation}} = \frac{3(37.46 - 37.1)}{\sqrt{0.077}} = 0.373$

$$\frac{\Sigma(x-\bar{x})^2}{n-1}$$

14

$Q_1 - 1.5\ IQR$ $\quad Q_3 + 1.5\ IQR$

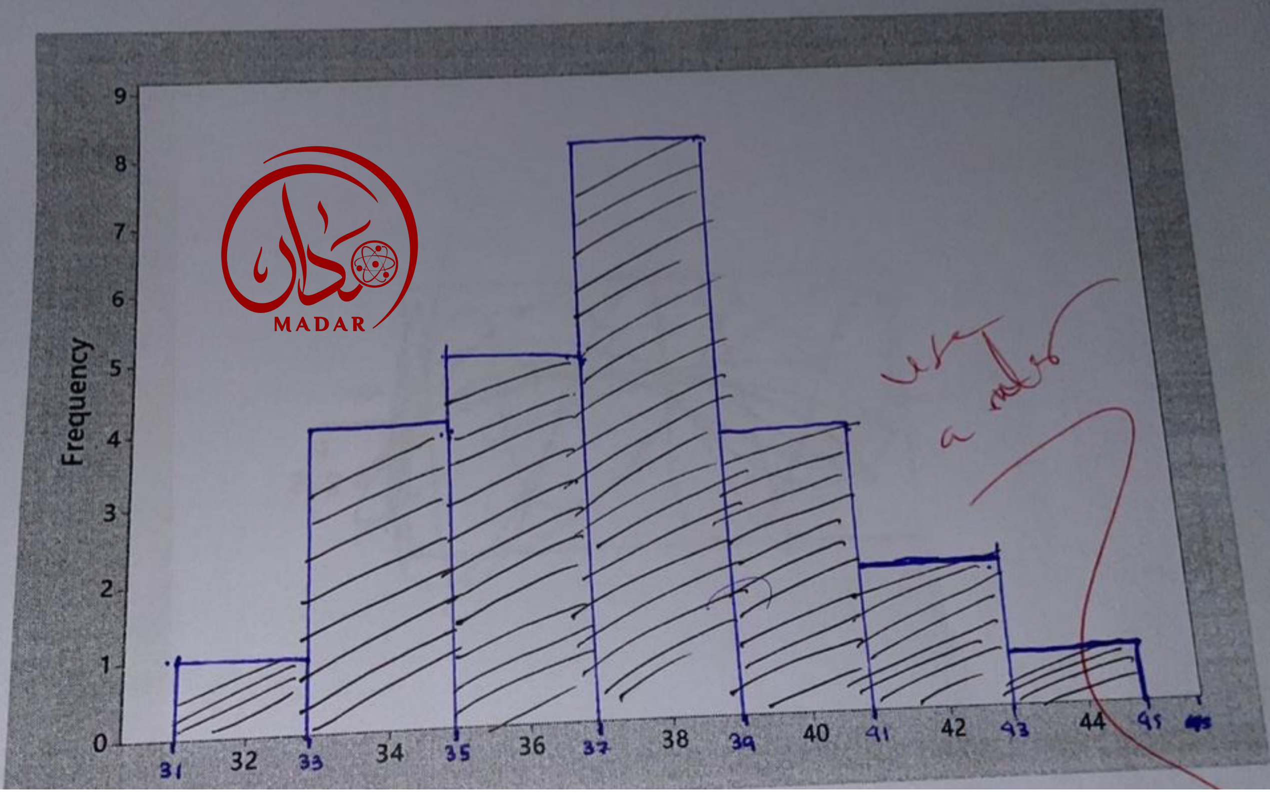
$$\left(\begin{matrix} \cancel{41.45} \\ \cancel{34.05} \\ 30.35 \end{matrix} - 45.15\right)$$

MADAR

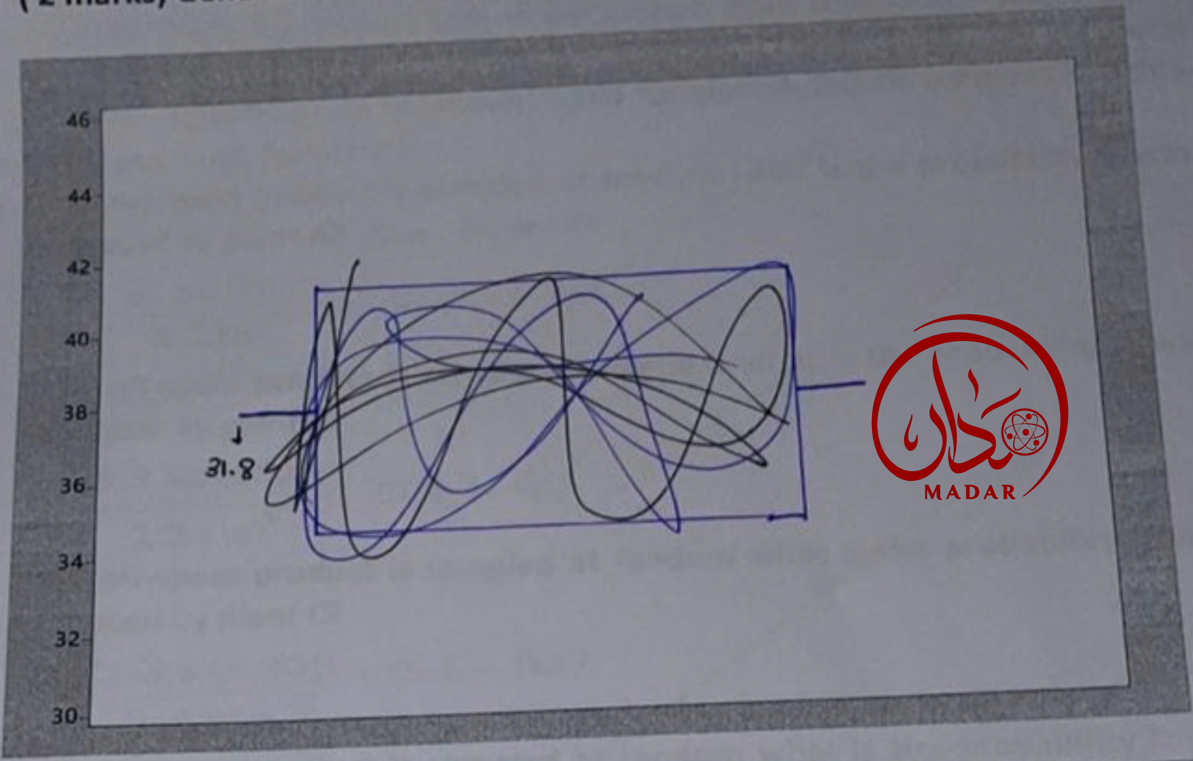
g) (4 marks) Generate a frequency table as below i.e., with seven intervals of width 2 for each interval beginning from 31 and ending at 45.

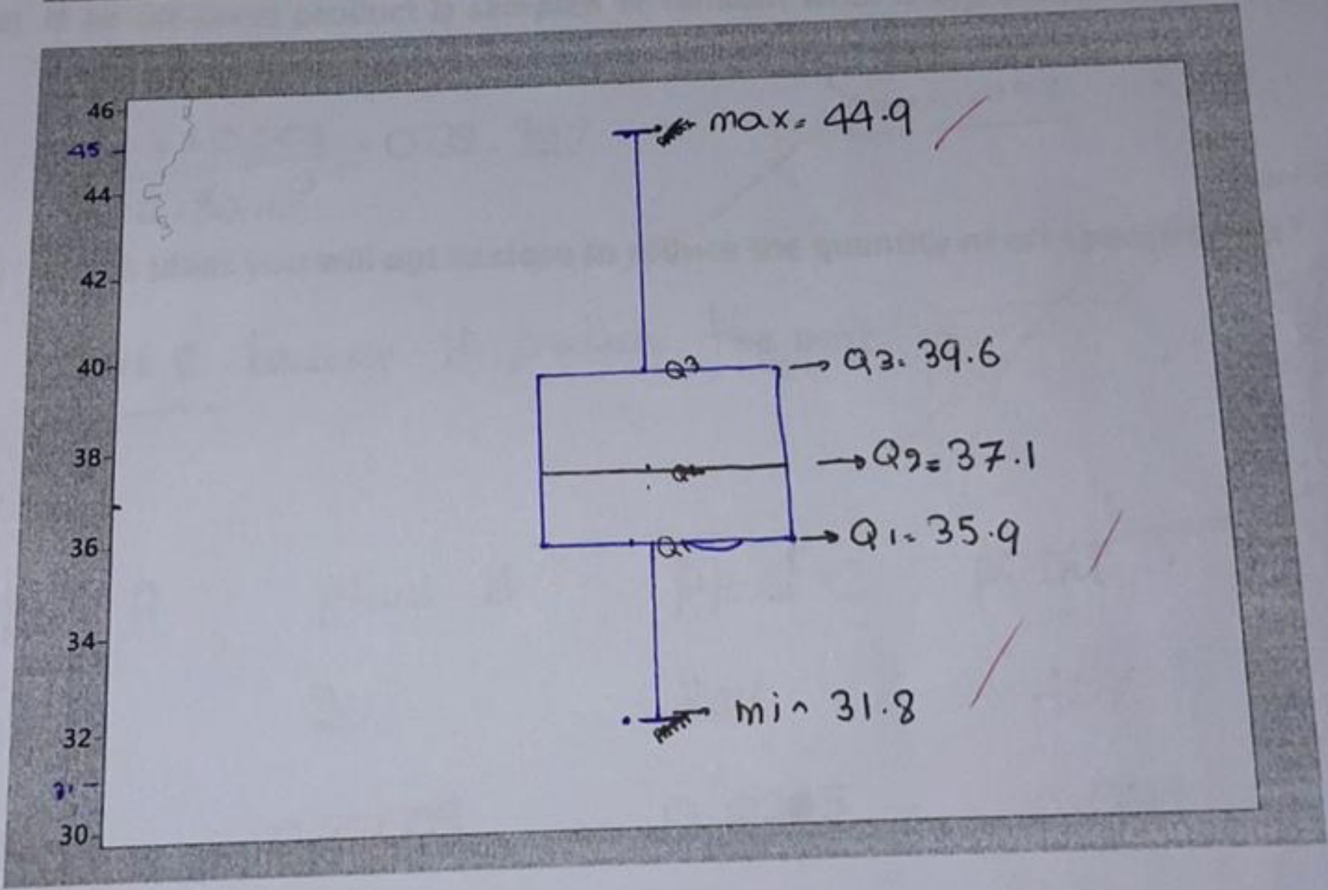
Mileage	Frequency (No. cars)	Relative frequency (%) fequancy/total	Cumulative frequency (%)
[31, 33)	1	0.04	0.04
[33, 35)	4	0.16	0.2
[35, 37)	5	0.2	0.4
[37, 39)	8	0.32	0.72
[39, 41)	4	0.16	0.88
[41, 43)	2	0.08	0.96
[43, 45)	1	0.04	1
Sum	25	1	

h) (3 marks) Generate a histogram of these data.



i) (2 marks) Generate a Boxplot of these data.





(30.35 - 45.15)

3. (8 Marks): Let X be a discrete random variable with the following probability mass function (PMF)

x	0.2	0.4	0.5	0.8	1.0
$f(x)$	0.1	0.2	0.2	0.3	0.2

a) (1 Mark) What is $P(X \le 0.5)$?

$f(0.2) + f(0.4) + f(0.5) = 0.1 + 0.2 + 0.2 = 0.5$

b) (1 Marks) What is $P(0.25 < X < 0.75)$?

$f(0.4) + f(0.5) = 0.2 + 0.2 = 0.4$

c) (2 Marks) Find $P(X = 0.2 \mid X < 0.6)$?

$= \dfrac{P(X=0.2 \cap X<0.6)}{P(X<0.6)} = \dfrac{0.1}{0.1+0.2+0.2} = 0.2$

$f(0.2)+f(0.4)+f(0.5)$

d) (2 Marks) What is E(X)?

$E(x) = 0.2\times0.1 + 0.4\times0.2 + 0.5\times0.2 + 0.8\times0.3 + 1\times0.2 = 0.64$

e) (2 Marks) What is Var(x)?

$\sum x^2 f(x) - \mu^2$

$0.2^2(0.1) + 0.4^2(0.2) + 0.5^2(0.2) + 0.8^2(0.3) + 1^2(0.2)$

$\frac{1}{250} + 0.032 + 0.05 + 0.192 + 0.2 =$

$0.478 - 0.64^2 = 0.0684$

8

2. (7 Marks) in a chemical plant there are four lines of production to produce the same product. Plant A produces 10% of the output, plant B produces 20%, plant C produces 30%, and plant D produces 40% of the output. The proportion of off-specs products from these lines are as follows: 0.001 for plant A, 0.0005 for plant B, 0.005 for plant C, and 0.002 for plant D.

a) If an off-specs product is sampled at random what is the probability it was produced by plant A? from bayes rule

$$\frac{0.001 \times 0.1}{2.5 \times 10^{-3}} = 0.04 = 4\%$$

b) If an off-specs product is sampled at random what is the probability it was produced by plant B?

$$\frac{0.2 \times 0.0005}{2.5 \times 10^{-3}} = 0.04 = 4\%$$

c) If an off-specs product is sampled at random what is the probability it was produced by plant C?

$$\frac{0.3 \times 0.005}{2.5 \times 10^{-3}} = 0.6 = 60\%$$

d) If an off-specs product is sampled at random what is the probability it was produced by plant D?

$$\frac{0.4 \times 0.002}{2.5 \times 10^{-3}} = 0.32 = 32\%$$

Sum = 100%

e) Which plant you will opt to close to reduce the quantity of off-spec product?

Plant C, because it produces the most

7

Plant A	Plant B	Plant C	Plant D
10%	20%	30%	40%
~~0.01~~ 0.001	0.0005	0.0005	0.002

now from total probability rule

$0.001 \times 0.1 + 0.2 \times 0.0005 + 0.3 \times 0.005 + 0.4 \times 0.002 = 2.5$

2. (7 Marks) in a chemical plant there are four lines of production to produce the same product. Plant A produces 10% of the output, plant B produces 20%, plant C produces 30%, and plant D produces 40% of the output. The proportion of off-specs products from these lines are as follows: 0.001 for plant A, 0.0005 for plant B, 0.005 for plant C, and 0.002 for plant D.

a) If an off-specs product is sampled at random what is the probability it was produced by plant A?

$$P(A|O) = \frac{P(O|A)\,P(A)}{P(O)} = \frac{0.1 \times 0.001}{0.0025} = 0.04$$

b) If an off-specs product is sampled at random what is the probability it was produced by plant B?

$$P(B|O) = \frac{P(O|B)\,P(B)}{P(O)} = \frac{0.2 \times 0.0005}{0.0025} = 0.04$$

c) If an off-specs product is sampled at random what is the probability it was produced by plant C?

$$P(C|O) = \frac{P(O|C)\,P(C)}{P(O)} = \frac{0.3 \times 0.005}{0.0025} = 0.6$$

d) If an off-specs product is sampled at random what is the probability it was produced by plant D?

$$P(D|O) = \frac{P(O|D)\,P(D)}{P(O)} = \frac{0.4 \times 0.002}{0.0025} = 0.32$$

e) Which plant you will opt to close to reduce the quantity of off-spec product?

[illegible] plant C because it has the ~~highest~~ Probability to produce an off spec Product.

7

plant	produce	proportion
A	10%	0.001
B	20%	0.0005
C	30%	0.005
D	40%	0.002

using bayes theorem

$$P(O) = (0.1 \times 0.001) + (0.2 \times 0.0005) + (0.3 \times 0.005) + (0.4 \times 0.002) = 0.0025$$

1. (6 Marks divided equally between all parts) a sample of surface temperature at various places on a certain day are given as (°C): 13.5, 14.0, 14.5, 15.0, 15.0, 15.0, 16.0, 16.5, and 16.5.

a) What is the mean of these data?

$\bar{x} = \dfrac{13.5 + 14 + 14.5 + 15 + 15 + 15 + 16 + \cancel{16 + 16} \; 16.5 + 16.5}{9} = 15.1$

b) What is the variance of these data? $6^2 = 0.987$. $6^2 = \dfrac{(x-\bar{x})}{N}$

$6^2 = \dfrac{(x-\bar{x})^2}{N} = 0.987$

c) What is the standard deviation of these data? $\sqrt{6^2}$

$S = \sqrt{6^2} = 0.9938$

d) What is the range of these data?

$13.5 < x < 16.5$

e) What is the median of these data?

15

f) What is the mode of these data?

3

$(13.5 - 15.1)^2 = 2$

2. (7.5 Marks) there are 100 candidate raw material to be used in producing a certain desired product. If you were to select five raw materials at random, what is the chance that the selection would contain exactly one of the preferred five raw materials?

$$P(A) = \frac{C^{5}_{1} \cdot C^{95}_{4}}{C^{100}_{5}} = \frac{5 * 3183545}{75287520} = 0.211425$$

details

6.5