	A	B	C	D	E	F	G
	23	33	46	129	33	321	461
	39	49	78	156	41	319	782
	32	42	64	145	49	231	643
	66	76	132	160	56	265	132
	58	68	116	119	85	541	126
	42	52	84	134	24	442	184
	37	47	74	170	73	358	274
	49	59	98	98	94	149	398
	47	57	94	144	74	333	394
	32	42	64	135	23	301	464
				162	82	329	156
				152	44	149	288
				147		231	
				136		149	
				152		333	
				138		256	
Mean							
Mode							
Median							
Variance							
St. Dev							
Max							
Min							
Range							
Est. ST. Dev							
Q1							
Q2							
Q3							
Box-Plot							
99%							
99%							
95%							
95%							
68%							
68%							

Questions A. Determine whether the given value is a **statistic** or a **parameter.**

1. A sample of divers is selected, and the average age is 41.8 years.

2. After checking computer records for every commercial movie made last year the longest running time is found to be 187 minutes

3. All of the cities mayors are surveyed, and the 250 out of them are found to be democrat.

4. The average speed of 35 drivers on the highway.

5. The average IQs of the top 10 Miss Universe finalists.

Questions B. Determine whether the given values are from a discrete or continuous data set.

1. A math teacher counts 4 absent students

2. The weight of SUVs made in Japan.

3. The annual average rainfall in California.

4. The times taken for athletes to run 100m.

5. The numbers of chocolates in various 500g boxes.

Questions C. Identify which type of sampling of sampling is used**:** **R**andom, **S**ystematic, **C**luster, **C**onvenience, or **S**tratified.

1. At a local Mall a researcher ask every 10^{th} passing by shoppers about the new security measure.

2. CNN is planning an exit poll in which 100 polling stations will be randomly selected and all voters will be interviewed as they leave the premises.

3. An engineering student measures the strengths of finger used to push buttons by testing family members.

4. An IRS researcher investigates cheating on income tax reports by surveying all waiters and waitresses at 20 randomly selected restaurants.

5. A marketing expert for MTV is planning a survey in which 500 people will be randomly selected from each age groups of 10-19, 20-29, and so on.

6. The author surveyed all of his students to obtain sample data consisting of the number of credit cards students possesses.

7. Fund-raisers for the college of Newport test a new telemarketing campaign by obtaining a list of all alumni and selecting every 1000^{th} name on the list.

8. In a Gallup poll of 1045 adults, the interview subjects were selected by using a computer to randomly generate telephone numbers that were then called.

9. A market researcher has partitioned all Californian residents into categories of unemployed, employed full time, and employed part time. She is surveying 50 people from each category.

10. Motivated by a student who died from binge drinking. The College of South land conducts a study of students drinking by randomly selecting 10 different classes and interviewing all students in each of those classes.

Question D.

There were six different stat classes that were offered last semester; one student was randomly selected from each class with his/her final score and the class average and standard deviation.

Joe got score of 83 when the class average was 71 with standard deviation of 6.5.
Moe got score of 88 when the class average was 76 with standard deviation of 7.5.
Nielo got score of 77 when the class average was 72 with standard deviation of 2.3.
April got score of 82 when the class average was 72 with standard deviation of 5.5.
Max got score of 82 when the class average was 71 with standard deviation of 5.
Alex got score of 82 when the class average was 72 with standard deviation of 6.

Use the Z-score formula to answer the following questions:

a)Who did relatively better in the class than the rest?
b)Who did relatively worse in the class than the rest?
c)Who got scores that can be considered unusual?

Answers

Questions A.
1. statistic**,** **2.** parameter, **3**. parameter, **4**. statistic, **5**. statistic

Questions B.

1. Discrete**,** **2.** Continuous, **3**. Continuous **4**. Continuous, **5**. Discrete

---

Questions C.

1. Systematic **2.** Cluster. **3.** Convenience. **4.** Cluster. **5**. Stratified

6. Convenience. **7.** Systematic. **8**. Random. 9. Stratified **10**. Cluster

---

Questions D.

Name	x	$\overline{x}$	**s**	$z = \dfrac{x - \overline{x}}{s}$
Joe	83	71	6.5	1.85
Moe	88	76	7.5	1.60
Nielo	77	72	2.3	2.17
April	82	72	5.5	1.82
Max	82	71	5	2.20
Alex	82	72	6	1.67

a)Max **b)** Moe **c)** Nielo and Max

A. **Grouped Data**

Age(Month)	f	m	Rel f %	$f \times m$	$f \times m^2$
1 - 3	5	2	10		
3 - 5	10			40	
5 - 7	15		30		540
7 - 9	12	8		96	
9 - 11	6		12		600
11 - 13	2	12			
	$n = \sum f =$		Add to 100%?	$\sum(f \times m) =$	$\sum(f \times m^2) =$

Draw the

1. Histogram (write your observation)
2. Frequency polygon

Compute.

3. Mean?(**6.4**)
4. Variance?(**6.69**)
5. Standard deviation? ***(2.6)***
6. Apply all three empirical rules.

$0 <$ **99.7 %** of data < 14.2 , $1.2 <$ **95 %** of data < 11.6 , $3.8 <$ **68 %** of data < 9

B.

Scores	f	m	Rel f %	$f \times m$	$f \times m^2$
00-10	**2**	**5**			
10-20	**6**		**6**		**1350**
20-30	**8**			**200**	**5000**
30-40	**14**			**490**	**17150**
40-50	**16**				**32400**
50-60	**14**	**55**			
60-70	**16**			**1040**	
70-80	**12**				
80-90	**8**		**8**		
90-100	**4**				
	$n = \sum f =$		Add to 100%?	$\sum(f \times m) =$	$\sum(f \times m^2) =$

Draw the

1. Histogram (write your observation)
2. Frequency polygon

Compute.

3. Mean? (***52.80)***
4. Variance? (***490.06)***
5. Standard deviation? (***22.14)***
6. Apply all three empirical rules.

$0 <$ **99.7 %** of class < 119.22 , $8.52 <$ **95 %** of class < 97.08 , $30.66 <$ **68 %** of class < 74.94

C.

Weights	f	m	Rel f %	$f \times m$	$f \times m^2$
25 - 35	1				
35 - 45	3				4800
45 - 55	7			350	
55 - 65	10				
65 - 75	11				
75 - 85	15				
85 - 95	18				
95 - 105	28		22.4		280000
105 - 115	32			3520	
	$n = \sum f =$		Add to 100%?	$\sum(f \times m) =$	$\sum(f \times m^2) =$

Draw the

1. Histogram (write your observation)
2. Frequency polygon

Compute.

3 Mean? ***(88.08)***

4. Variance? ***(422.09)***

5. Standard deviation? ***(20.54)***

6. Apply all three empirical rules.

D.

Time(sec)	f	m	Rel f %	$f \times m$	$f \times m^2$
6 - 12	**100**		40		
12 - 18	**60**	**15**			
18 - 24	**50**				**22050**
24 - 30	**20**		8		
30 - 36	**8**	**33**			
36 - 42	**6**		2.4		
42 - 48	**4**				**8100**
48 - 54	**2**		0.8	**102**	
	$n = \sum f =$		Add to 100%?	$\sum(f \times m) =$	$\sum(f \times m^2) =$

Draw the

1. Histogram (write your observation)
2. Frequency polygon

Compute.

3. Mean? (***16.68)***

4. Variance? (***79.58)***

5. Standard deviation?***(8.92)***

6. Apply all three empirical rules.

Abe Mirza **Regression and correlation**

A.

	x = Hours Study/week	y = Test Score	x^2	y^2	$x\,y$
1 5		72	25	5184	360
2 10		88	100	7764	880
3 13		92	169	8464	1196
4 8		80	64	6400	640
5 6		77	36	5926	462
6 4		64	16	4096	256
	$\sum x = 46$	$\sum y = 473$	$\sum x^2 = 410$	$\sum y^2 = 37817$	$\sum xy = 3794$

1. Use the data and plot the data as a scattered diagram and **comment** on the pattern of the points.
2. Compute the correlation coefficient and **comment** on that $r = 0.963$ Very strong…?
3. Compute the slope and y-intercept and write the equation of regression line. Slope = a = 2.92, y-itc = b = 56.41
   y = a x + b = 2.92 x + 56.41
4. Explain the slope based on the regression equation and the in relation of x and y variables.
   In general for every additional hour of study per week the score goes up by 2.92 points.
5. Compute average and standard deviation for both x and y variables. $\bar{x} = 7.67$, $\bar{y} = 78.83$, $s_x = 3.386$, $s_y = 10.28$
6. If one student studies 7 hours a week, use **Reg. Equ.** to estimate her test score. $x = 7$, $y' = 73.93$
7. If one student has test score of 85, use **Reg. Equ.** to estimate number of hours he spends studying per week.
   $y = 85$, $x' = 9.79$
8. Compute the coefficient of determination $(r^2 \times 100)$ and **comment** on that.
   $(r^2 \times 100) = (.962^2 \times 100) = 92\ \%$, *92% of variations in test score are explained by regression equation.*

B

X = Experience(yrs)	14	3	5	6	4	9	18	5	16
Y = Monthly Salary $(000)	42	24	33	31	29	39	47	30	43

1. Use the data and plot the data as a scattered diagram and **comment** on the pattern of the points.

2. Compute the correlation coefficient and **comment** on that__________________

3. Compute the slope and y-intercept and write the equation of regression line._____________________

4. E xplain the slope based on the regression equation and the in relation of x and y variables.

5. Compute average and standard deviation for both x and y variables._________________________

6. If some one's experience is 10 years old, use **Reg. Equ.** to estimate his salary. _________________

7. If some one's salary is $38, 000, use **Reg. Equ.** to estimate her experience. __________________

8. Compute the coefficient of determination and **comment** on that.

C

X = Year (1998=0)	0	1	2	3	4	5	6	7	8	9	10
Y= Net connected PCs(mil)	22	32	45	58	70	86	99	119	140	155	178

1. Use the data and plot the data as a scattered diagram and **<u>comment</u>** on the pattern of the points.

2. Compute the correlation coefficient and **<u>comment</u>** on that_________________

3. Compute the slope and y-intercept and write the equation of regression line.____________________

4. Explain the slope based on the regression equation and the in relation of x and y variables.

5. Compute average and standard deviation for both x and y variables.______________________

6. Use **Reg. Equ.** to estimate how many PCs will be connected by year 2009? _____________

7. Use **Reg. Equ.** to estimate in what year about 250 million PCs are net connected.___________

8. Compute the coefficient of determination and **<u>comment</u>** on that.

D

X= IQ Score	120	140	130	150	142	130	135	175	149	168
Y=Reading Score	62	62	63	65	66	67	68	68	70	72

1. Use the data and plot the data as a scattered diagram and **<u>comment</u>** on the pattern of the points.

2. Compute the correlation coefficient and **<u>comment</u>** on that_________________

3. Compute the slope and y-intercept and write the equation of regression line.____________________

4. Explain the slope based on the regression equation and the in relation of x and y variables.

5. Compute average and standard deviation for both x and y variables.______________________

6. If some one’s IQ score is 100 estimate her reading score.________________

7. If some one’s reading score is 86 estimate his IQ score. __________________

8. Compute the coefficient of determination and **<u>comment</u>** on that.

E.

X= Midterm	75	68	82	91	84	77	72	88	90	66	70	81	59
Y= Final	77	72	80	89	89	80	72	88	92	70	72	83	66

1. Use the data and plot the data as a scattered diagram and **comment** on the pattern of the points.

2. Compute the correlation coefficient and **comment** on that_________________

3. Compute the slope and y-intercept and write the equation of regression line.____________________

4. Explain the slope based on the regression equation and the in relation of x and y variables.

5. Compute average and standard deviation for both x and y variables.________________________

6. If some one gets 74 on the midterm estimate his final score._______________

7. If some one gets 74 on the final estimate her midterm score._______________

8. Compute the coefficient of determination and **comment** on that.

F.

X = Number of times absent	2	3	5	2	6	0	4	3	9	5	0	4	8
Y = Average test scores	92	88	80	85	71	85	74	77	65	70	89	76	67

1. Use the data and plot the data as a scattered diagram and **comment** on the pattern of the points.

2. Compute the correlation coefficient and **comment** on that_________________

3. Compute the slope and y-intercept and write the equation of regression line.____________________

4. Explain the slope based on the regression equation and the in relation of x and y variables.

5. Compute average and standard deviation for both x and y variables.________________________

6. If some one has been absent 7 times, then estimate his average test score.______________

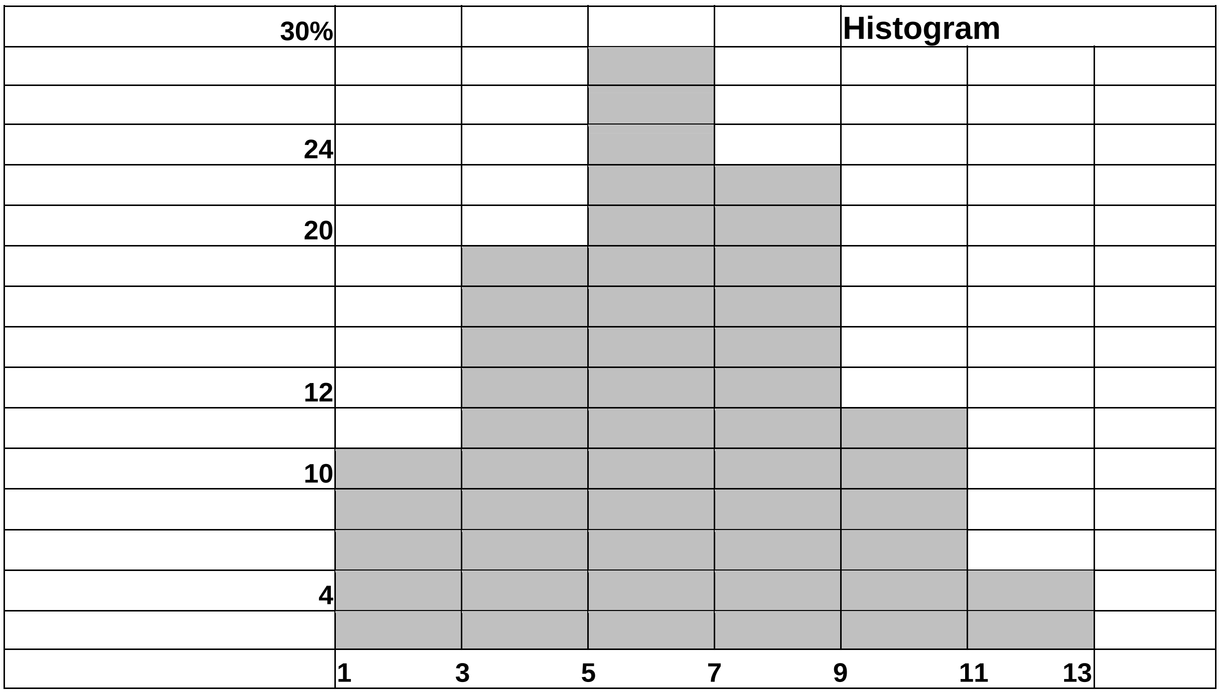
7. If some one's average test score is 90, then estimate the number of absentees she might have______

8. Compute the coefficient of determination and **comment** on that.

Group Data Answer Frequency Table

Age(Month)	f	m	Rel f %	$f \times m$	$f \times m^2$
1 - 3	5	2	10	10	20
3 - 5	10	4	20	40	160
5 - 7	15	6	30	90	540
7 - 9	12	8	24	96	768
9 - 11	6	10	12	60	600
11 - 13	2	12	4	24	288
	$n = \sum f = \mathbf{50}$		**100%**	$\sum (f \times m) = \mathbf{320}$	$\sum (f \times m^2) = \mathbf{2376}$

Problem A



Ages (Months)

3. **Mean**: $\bar{X} = \dfrac{\sum (f \times m)}{n} = \dfrac{320}{50} = \mathbf{6.4}$

4. **Variance**: $S^2 = \dfrac{n\sum (f \times m^2) - \left(\sum (f \times m)\right)^2}{n(n-1)} = \dfrac{50(\mathbf{2376}) - (\mathbf{320})^2}{50(50-1)} = \dfrac{16400}{2450} = \mathbf{6.69}$

5. **Standard deviation** = $S = \sqrt{\mathbf{6.69}} = \mathbf{2.59} = \mathbf{2.6}$

6.

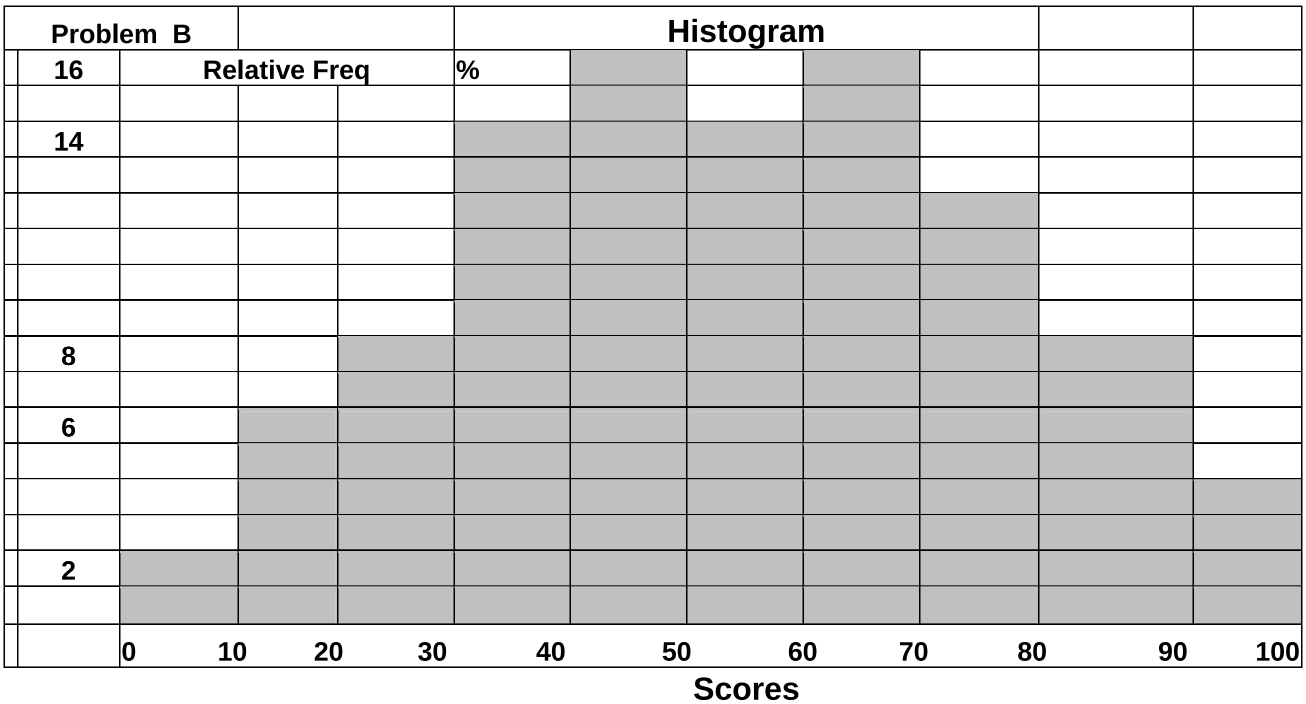
Histogram is centered so the results of empirical rules will be valid.

$99.7\% = 6.4 \pm 3(2.6) = 6.4 \pm 7.8$ $0 <$ **99.7 %** of data < 14.2

$95\% = 6.4 \pm 2(2.6) = 6.4 \pm 5.2$ $1.2 <$ **95 %** of data < 11.6

$68\% = 6.4 \pm 1(2.6) = 6.4 \pm 2.6$ $3.8 <$ **68 %** of data < 9

B					
Scores	**f**	**m**	**Rel f %**	$f \times m$	$f \times m^2$
00-10	**2**	**5**	**2**	**10**	**50**
10-20	**6**	**15**	**6**	**90**	**1350**
20-30	**8**	**25**	**8**	**200**	**5000**
30-40	**14**	**35**	**14**	**490**	**17150**
40-50	**16**	**45**	**16**	**720**	**32400**
50-60	**14**	**55**	**14**	**770**	**42350**
60-70	**16**	**65**	**16**	**1040**	**67600**
70-80	**12**	**75**	**12**	**900**	**67500**
80-90	**8**	**85**	**8**	**680**	**57800**
90-100	**4**	**95**	**4**	**380**	**36100**
	$n = \sum f = 100$		**100%**	$\sum (f \times m) = 5280$	$\sum (f \times m^2) = 327300$



3. **Mean**: $\overline{X} = \dfrac{\sum (f \times m)}{n} = \dfrac{5280}{100} = \mathbf{52.80}$

4. **Variance**: $S^2 = \dfrac{n \sum (f \times m^2) - \left(\sum (f \times m)\right)^2}{n(n-1)} = \dfrac{100(\mathbf{327300}) - (\mathbf{5280})^2}{100(100-1)} = \dfrac{4851600}{9900} = \mathbf{490.06}$

5. **Standard deviation** = $S = \sqrt{\mathbf{490.06}} = \mathbf{22.14}$

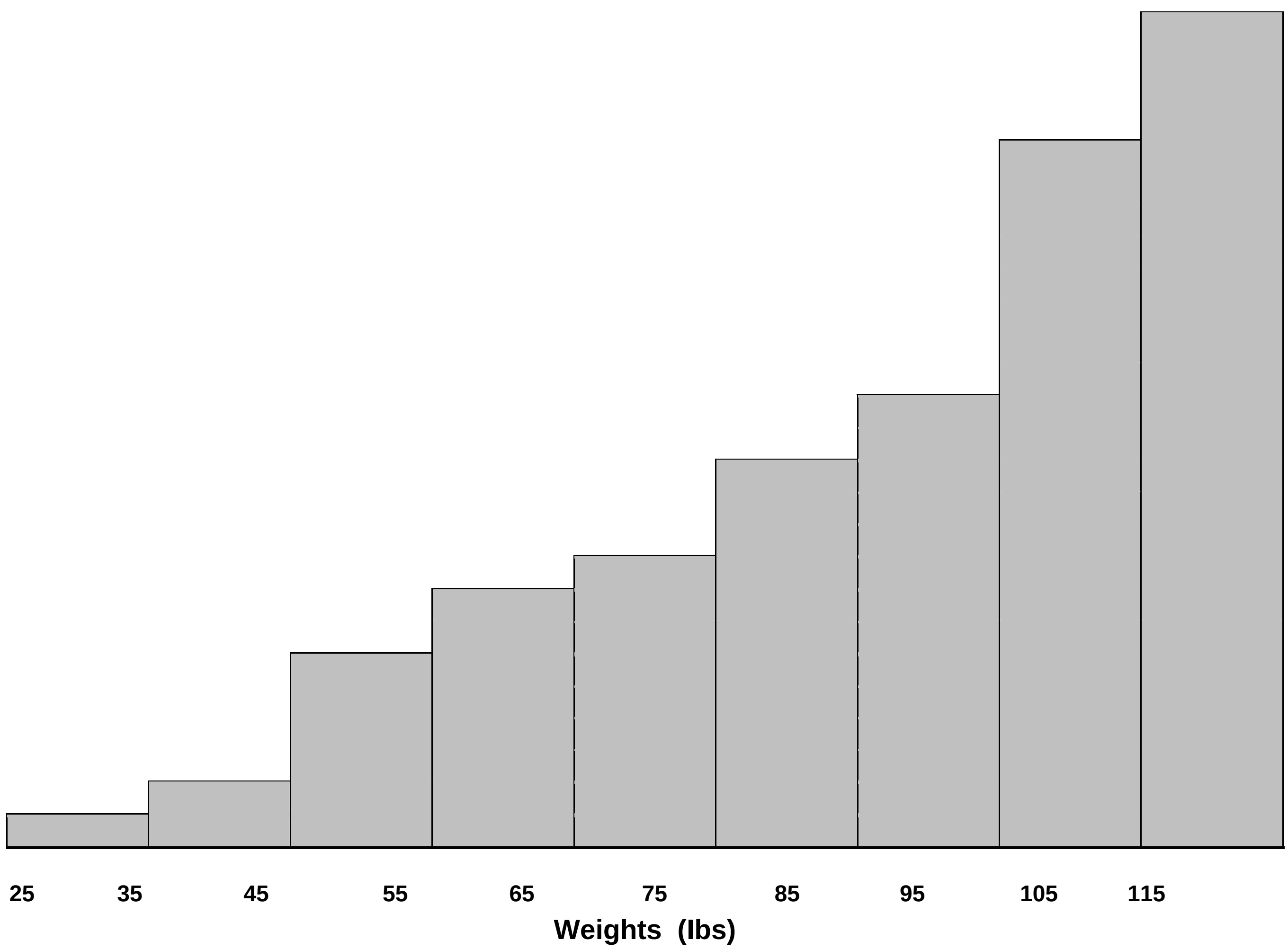
Histogram is relatively centered so the results of empirical rules will be valid.

$99.7\% = 52.8 \pm 3(22) = 52.8 \pm 66$ $\qquad 0 <$ **99.7 %** of class got scores < 118.8

$95\% = 52.8 \pm 2(22) = 52.8 \pm 44$ $\qquad 8.8 <$ **95 %** of class got scores < 96.8

$68\% = 52.8 \pm 1(22) = 52.8 \pm 22$ $\qquad 30.8 <$ **68 %** of class got scores < 74.8

Problem C					
Weights	**f**	**m**	**Rel f %**	$f \times m$	$f \times m^2$
25 - 35	1	30	.8	30	900
35 - 45	3	40	2.4	120	4800
45 - 55	7	50	5.6	350	17500
55 - 65	10	60	8	600	36000
65 - 75	11	70	8.8	770	53900
75 - 85	15	80	12	1200	96000
85 - 95	18	90	14.4	1620	145800
95 - 105	28	100	22.4	2800	280000
105 - 115	32	110	25.6	3520	387200
	$n = \sum f = \mathbf{125}$		100%	$\sum f \times m = \mathbf{11010}$	$\sum f \times m^2 = \mathbf{1022100}$



3. Mean: $\overline{X} = \frac{\sum (f \times m)}{n} = \frac{11010}{125} = \mathbf{88.08}$

4. Variance: $S^2 = \frac{n \sum (f \times m^2) - \left(\sum (f \times m)\right)^2}{n(n-1)} = \frac{125(\mathbf{1022100}) - (\mathbf{11010})^2}{125(125-1)} = \frac{6542400}{15500} = \mathbf{422.09}$

5. Standard deviation = $S = \sqrt{\mathbf{422.09}} = \mathbf{20.54}$

Histogram is not centered so the results of empirical rules will not be valid.

$99.7\% = 88.08 \pm 3(20.54) = 88.08 \pm 61.62$ — $26.46 <$ **99.7 %** of weights are between < 149.7

$95\% = 88.08 \pm 2(20.54) = 88.08 \pm 41.08$ — $47 <$ **95 %** of weights are between < 129.16

$68\% = 88.08 \pm 1(20.54) = 88.08 \pm 20.54$ — $67.54 <$ **68 %** of weights are between < 108.62

D					
Time(sec)	**f**	**m**	**Rel f %**	$f \times m$	$f \times m^2$
6 - 12	100	9	40	900	8100
12 - 18	60	15	24	900	13500
18 - 24	50	21	20	1050	22050
24 - 30	20	27	8	540	14580
30 - 36	8	33	3.2	264	8712
36 - 42	6	39	2.4	234	9126
42 - 48	4	45	1.6	180	8100
48 - 54	2	51	0.8	102	5202
	$n = \sum f = \mathbf{250}$		**100%**	$\sum(f \times m) = \mathbf{4170}$	$\sum(f \times m^2) = \mathbf{89370}$

3. **Mean**: $\bar{X} = \dfrac{\sum(f \times m)}{n} = \dfrac{4170}{250} = \mathbf{16.68}$

4. **Variance**: $S^2 = \dfrac{250(\mathbf{89370}) - (\mathbf{4170})^2}{250(250-1)} = \mathbf{79.58}$ 5. **Standard deviation** = $S = \sqrt{79.58} = \mathbf{8.92}$

Histogram is not centered so the results of empirical rules will not be valid.

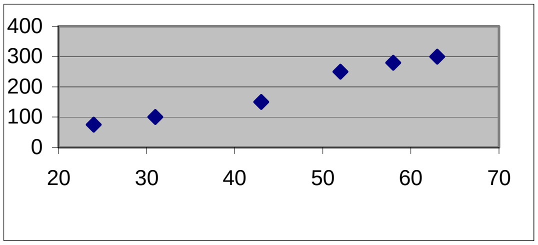
$99.7\% = 16.68 \pm 3(8.92) = 16.68 \pm 26.76$ — $0 <$ **99.7 %** of Times are between < 43.44

$95\% = 16.68 \pm 2(8.92) = 16.68 \pm 17.84$ — $0 <$ **95 %** of Times are between < 34.52

$68\% = 16.68 \pm 1(8.92) = 16.68 \pm 8.92$ — $7.76 <$ **68 %** of Times are between < 25.6

Answers

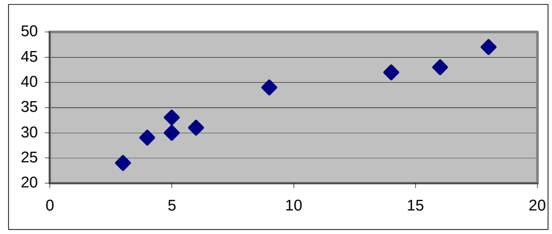
Problem A



	x	y
Mean	**7.67**	**78.83**
St Dev.	**3.386**	**10.28**
Correl Coeff	$r =$ **0.963**	
Slope	**2.92**	
Y-itc	**56.41**	

Y = 2.92 X + 56.41		
X = 6 ,	y' **= ? = 73.93**	
Y = 85 ,	x' **= ? = 9.79**	

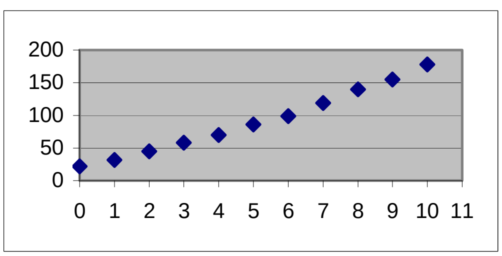
Problem B



	x	y
Mean	**8.889**	**35.333**
St Dev.	**5.667**	**7.697**
Correl Coeff	$r =$ **0.961**	
Slope	**1.305**	
Y-itc	**23.730**	

Y = 1.305 X + 23.73		
X = 10 ,	y' **= ? = 36.78**	
Y = 38 ,	x' **= ? = 10.93**	

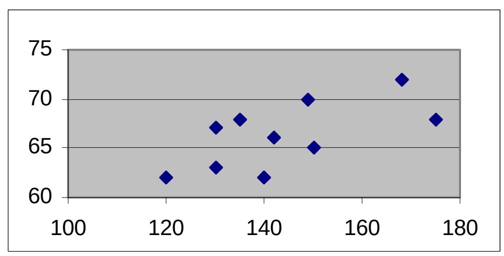
Problem C



	x	y
Mean	**5.000**	**91.273**
St Dev.	**3.317**	**51.794**
Correl Coeff	$r =$ **0.994**	
Slope	**15.527**	
Y-itc	**13.636**	

Y = 15.527 X +13.636		
X = 11 ,	y' **= ? =** 184.47	
Y = 250 ,	x' **= ? =** $15.22 = 2005$	

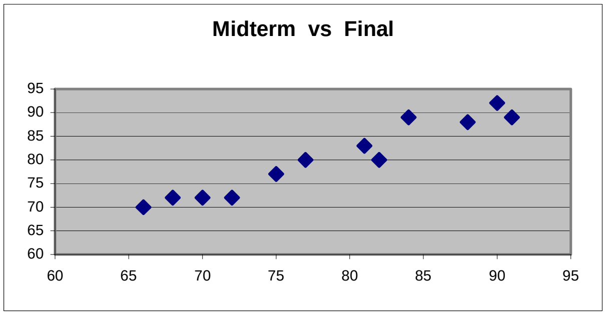
Problem D



	x	y
Mean	**143.900**	**66.300**
St Dev.	**17.214**	**3.368**
Correl Coeff	$r =$ **0.656**	
Slope	**0.128**	
Y-itc	**47.830**	

Y = 0.128 X + 47.83		
X = 100 ,	y' **= ? =** 60.63	
Y = 86 ,	x' **= ? =** 298.20	

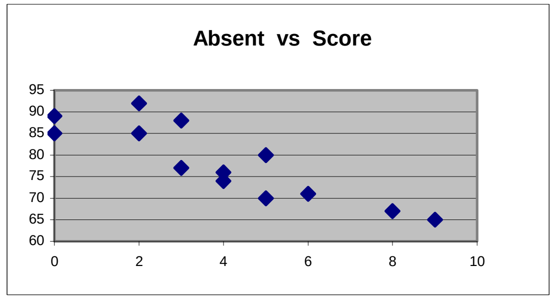
Problem E



Mean	77.154	79.231
St Dev.	9.915	8.506
Correl Coeff	$r = 0.971$	
Slope	0.833	
Y-itc	14.971	

Y = 0.833 X +14.971	
X = 74 ,	y' = ? = 76.61
Y = 74 ,	x' = ? = 70.86

Problem F



Mean	3.923	78.385
St Dev.	2.722	8.856
Correl Coeff	$r = -0.870$	
Slope	-2.830	
Y-itc	89.485	

Y = $-$2.83 X + 89.485	
X = 7 ,	y' = ? = 69.68
Y = 90 ,	x' = ? = -0.18

Answers to Ungrouped Data

	A	B	C	D	E	F	G
	23	33	46	129	33	321	461
	39	49	78	156	41	319	782
	32	42	64	145	49	231	643
	66	76	132	160	56	265	132
	58	68	116	119	85	541	126
	42	52	84	134	24	442	184
	37	47	74	170	73	358	274
	49	59	98	98	94	149	398
	47	57	94	144	74	333	394
	32	42	64	135	23	301	464
				162	82	329	156
				152	44	149	288
				147		231	
				136		149	
				152		333	
				138		256	
Mean	***42.50***	***52.50***	***85.00***	***142.31***	***56.50***	***294.19***	***358.50***
Mode	***32***	***42***	***64***	***152***		***149***	
Median	***40.5***	***50.5***	***81***	***144.5***	***52.5***	***310***	***341***
Variance	***166.50***	***166.50***	***666.00***	***315.96***	***602.82***	***11065.10***	***43061.36***
St. Dev	***12.90***	***12.90***	***25.81***	***17.78***	***24.55***	***105.19***	***207.51***
Max	***66***	***76***	***132***	***170***	***94***	***541***	***782***
Min	***23***	***33***	***46***	***98***	***23***	***149***	***126***
Range	***43***	***43***	***86***	***72***	***71***	***392***	***656***
Est St. Dev	***10.75***	***10.75***	***21.5***	***18***	***17.75***	***98***	***164***
Q1	***32***	***42***	***64***	***134***	***37***	***231***	***170***
Q2	***40.5***	***50.5***	***81***	***144.5***	***52.5***	***310***	***341***
Q3	***49***	***59***	***98***	***154***	***78***	***333***	***462.5***
Box_Plot							
99%	***81.21***	***91.21***	***162.42***	***195.64***	***130.16***	***609.76***	***981.04***
99%	***3.79***	***13.79***	***7.58***	***88.99***	***-17.16***	***-21.38***	***-264.04***
95%	***68.31***	***78.31***	***136.61***	***177.86***	***105.60***	***504.57***	***773.52***
95%	***16.69***	***26.69***	***33.39***	***106.76***	***7.40***	***83.81***	***-56.52***
68%	***55.40***	***65.40***	***110.81***	***160.09***	***81.05***	***399.38***	***566.01***
68%	***29.60***	***39.60***	***59.19***	***124.54***	***31.95***	***189.00***	***150.99***