Tutorial 1

Inferential Statistics, Statistical Modelling & Survey Methods

(BS2506)

Pairach Piboonrungroj
(Champ)
me@pairach.com

1.(a) Pearson's product moment correlation coefficient and test a significant positive correlation

Month	FTSE100, X	Company, Y
Jan	4.64	6.11
Feb	2.36	3.84
Mar	0.98	2.07
Apr	1.82	-0.89
May	-0.41	2.98
Jun	0.9	3.53
Jul	9.57	10.69
Aug	1.53	-3.82
Sep	10.3	8.63
Oct	5.12	6.14
Nov	5.73	1.5
Dec	-0.27	1.24
Sum	**42.27**	**42.02**

$$r = \frac{\Sigma xy - n\bar{x}\bar{y}}{\sqrt{(\Sigma x^2 - n\bar{x}^2)(\Sigma y^2 - n\bar{y}^2)}}$$

1.(a) Pearson's product moment correlation coefficient and test a significant positive correlation

Month	FTSE100, X	Company, Y	XY
Jan	4.64	6.11	28.35
Feb	2.36	3.84	9.06
Mar	0.98	2.07	2.03
Apr	1.82	-0.89	-1.62
May	-0.41	2.98	-1.22
Jun	0.9	3.53	3.18
Jul	9.57	10.69	102.30
Aug	1.53	-3.82	-5.84
Sep	10.3	8.63	88.89
Oct	5.12	6.14	31.44
Nov	5.73	1.5	8.60
Dec	-0.27	1.24	-0.33
Sum	**42.27**	**42.02**	**264.82**

1.(a) Pearson's product moment correlation coefficient and test a significant positive correlation

Month	FTSE100, X	Company, Y	XY	X^2
Jan	4.64	6.11	28.35	21.53
Feb	2.36	3.84	9.06	5.57
Mar	0.98	2.07	2.03	0.96
Apr	1.82	-0.89	-1.62	3.31
May	-0.41	2.98	-1.22	0.17
Jun	0.9	3.53	3.18	0.81
Jul	9.57	10.69	102.30	91.58
Aug	1.53	-3.82	-5.84	2.34
Sep	10.3	8.63	88.89	106.09
Oct	5.12	6.14	31.44	26.21
Nov	5.73	1.5	8.60	32.83
Dec	-0.27	1.24	-0.33	0.07
Sum	**42.27**	**42.02**	**264.82**	**291.49**

1.(a) Pearson's product moment correlation coefficient and test a significant positive correlation

Month	FTSE100, X	Company, Y	XY	X^2	Y^2
Jan	4.64	6.11	28.35	21.53	37.33
Feb	2.36	3.84	9.06	5.57	14.75
Mar	0.98	2.07	2.03	0.96	4.28
Apr	1.82	-0.89	-1.62	3.31	0.79
May	-0.41	2.98	-1.22	0.17	8.88
Jun	0.9	3.53	3.18	0.81	12.46
Jul	9.57	10.69	102.30	91.58	114.28
Aug	1.53	-3.82	-5.84	2.34	14.59
Sep	10.3	8.63	88.89	106.09	74.48
Oct	5.12	6.14	31.44	26.21	37.70
Nov	5.73	1.5	8.60	32.83	2.25
Dec	-0.27	1.24	-0.33	0.07	1.54
Sum	**42.27**	**42.02**	**264.82**	**291.49**	**323.33**

1.a

$$r = \frac{\Sigma xy - n\bar{x}\bar{y}}{\sqrt{(\Sigma x^2 - n\bar{x}^2)(\Sigma y^2 - n\bar{y}^2)}}$$

$$r = \frac{264.8 - 12(3.52)(3.50)}{\sqrt{(291.48 - 12 * 12.39)(323.28 - 12 * 12.25)}} = 0.74$$

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{cases}$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$t = \frac{0.74\sqrt{10}}{\sqrt{1-0.563}} = 3.54$$

$$t = 3.54 > t_{10,0.05} = 1.812$$

thus we reject H_0

1(b)

$$\hat{\beta}_1 = \frac{\sum xy - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$$

$$\hat{\beta}_1 = \frac{264.8 - 12(3.52)(3.50)}{291.49 - (12)(3.52)^2} = \frac{116.96}{142.3} = 0.82$$

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

$$\hat{\beta}_0 = 3.50 - (0.82)(3.52) = 0.62$$

$$\hat{y} = 0.62 + 0.82x$$

1(c)

$$\begin{cases} H_0 : \beta_1 = 1 \\ H1 : \beta_1 < 1 \end{cases}$$

$$t = \frac{\hat{\beta}_1}{SE(\hat{\beta})}$$

$$SE(\hat{\beta}_1) = \frac{S}{\sqrt{\Sigma x^2 - n\bar{x}}} = \frac{2.837}{\sqrt{291.49 - 12(3.52)^2}} = 0.238$$

$$t = \frac{\hat{\beta}_1}{SE(\hat{\beta})} = \frac{0.82 - 1.0}{0.238} = -0.756$$

$$t_{0.05,10} = 1.812$$

$$t = -0.756 > t_c(-1.812)$$

Therefore do not reject the null hypothesis

2(a)

Country	Life Expectancy	Per-capita GNP
Algeria	64	2170
Botswana	59	940
Egypt	57	630
Ghana	55	380
Liberia	54	450
Libya	[illegible]	[illegible]
Malawi	[illegible]	[illegible]
Morocco	[illegible]	[illegible]
Senegal	[illegible]	[illegible]
South Africa	[illegible]	[illegible]
Sudan	[illegible]	[illegible]
Togo	55	390
Tunisia	64	1260
Uganda	49	250
Zimbabwe	60	640
Sum		

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

2(a)

Country	Life Expectancy	Rank	Per-capita GNP
Algeria	64	13	2170
Botswana	59	9	940
Egypt	57	8	630
Ghana	55	6.5	380
Liberia	54	5	450
Libya	67	15	5410
Malawi	49	2.5	180
Morocco	62	11	900
Senegal	47	1	650
South Africa	64	13	2460
Sudan	51	4	420
Togo	55	6.5	390
Tunisia	64	13	1260
Uganda	49	2.5	250
Zimbabwe	60	10	640
Sum			

2(a)

Country	Life Expectancy	Rank	Per-capita GNP	Rank
Algeria	64	13	2170	13
Botswana	59	9	940	11
Egypt	57	8	630	7
Ghana	55	6.5	380	3
Liberia	54	5	450	6
Libya	67	15	5410	15
Malawi	49	2.5	180	1
Morocco	62	11	900	10
Senegal	47	1	650	9
South Africa	64	13	2460	14
Sudan	51	4	420	5
Togo	55	6.5	390	4
Tunisia	64	13	1260	12
Uganda	49	2.5	250	2
Zimbabwe	60	10	640	8
Sum				

2(a)

Country	Life Expectancy	Rank	Per-capita GNP	Rank	d^2
Algeria	64	13	2170	13	0
Botswana	59	9	940	11	4
Egypt	57	8	630	7	1
Ghana	55	6.5	380	3	12.25
Liberia	54	5	450	6	1
Libya	67	15	5410	15	0
Malawi	49	2.5	180	1	2.25
Morocco	62	11	900	10	1
Senegal	47	1	650	9	64
South Africa	64	13	2460	14	1
Sudan	51	4	420	5	1
Togo	55	6.5	390	4	6.25
Tunisia	64	13	1260	12	1
Uganda	49	2.5	250	2	0.25
Zimbabwe	60	10	640	8	4
Sum					**99**

2(a)

Country	Life Expectancy	Rank	Per-capita GNP	Rank	d^2
Algeria	64	13	2170	13	0
Botswana	59	9	940	11	4
Egypt	57	8	630	7	1
Ghana	55	6.5	380	3	12.25
Liberia	54	5	450	6	1
Libya	67	15	5410	15	0
Malawi	49	2.5	180	1	2.25
Morocco	62	11	900	10	1
Senegal	47	1	650	9	64
South Africa	64	13	2460	14	1
Sudan	51	4	420	5	1
Togo	55	6.5	390	4	6.25
Tunisia	64	13	1260	12	1
Uganda	49	2.5	250	2	0.25
Zimbabwe	60	10	640	8	4
Sum					**99**

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6 * 99}{15(225 - 1)} = 0.82$$

2(b)

$$\begin{cases} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{cases}$$

$$Z = r\sqrt{(n-1)}$$

$$= 0.82\sqrt{14}$$

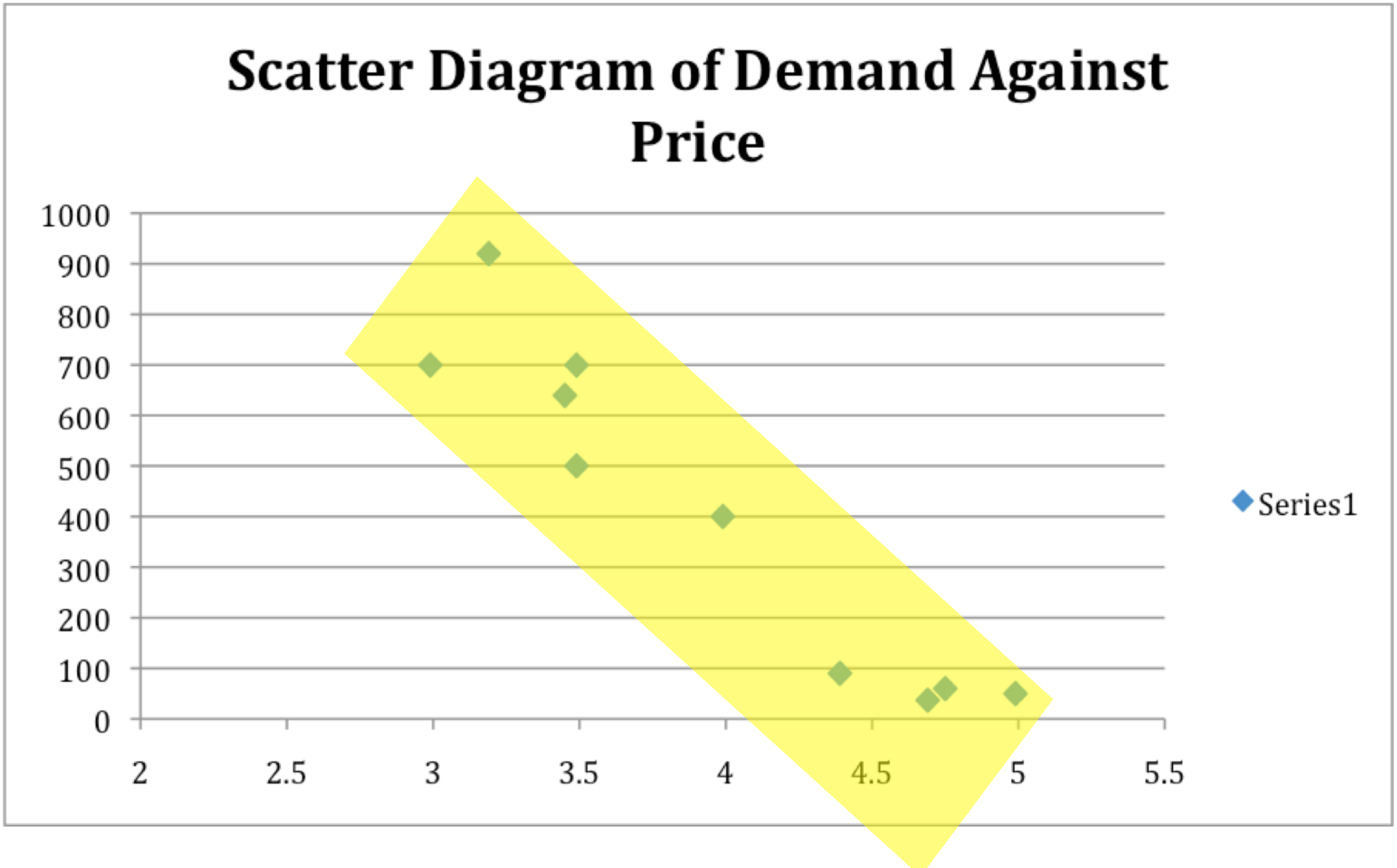
$$= 3.068 > Z_{(0.005 = 0.01/2)}(2.5758)$$

Reject H_0. Therefore there is a significant correlation between the rankings.

3. Wine’s Demand & Price

Sample	**Price (x)**	**Demand (y)**
1	3.99	400
2	3.45	640
3	2.99	700
4	3.49	700
5	4.69	37
6	4.99	50
7	3.49	500
8	4.75	60
9	4.39	90
10	3.19	920

3 (a)



Scatter Diagram of Demand Against Price
1000
900
800
700
600
500
400
300
200
100
0
2
2.5
3
3.5
4
4.5
5
5.5
Series1

3.b. Regression Model

Sample	Price (x)	Demand (y)
1	3.99	400
2	3.45	640
3	2.99	700
4	3.49	700
5	4.69	37
6	4.99	50
7	3.49	500
8	4.75	60
9	4.39	90
10	3.19	920

3(b)

Sample	Price (x)	Demand (y)	xy	x^2	y^2
1	3.99	400	1596	15.9201	160000
2	3.45	640	2208	11.9025	409600
3	2.99	700	2093	8.9401	490000
4	3.49	700	2443	12.1801	490000
5	4.69	37	173.53	21.9961	1369
6	4.99	50	249.5	24.9001	2500
7	3.49	500	1745	12.1801	250000
8	4.75	60	285	22.5625	3600
9	4.39	90	395.1	19.2721	8100
10	3.19	920	2934.8	10.1761	846400

3(b)

Sample	Price (x)	Demand (y)	xy	x^2	y^2
1	3.99	400	1596	15.9201	160000
2	3.45	640	2208	11.9025	409600
3	2.99	700	2093	8.9401	490000
4	3.49	700	2443	12.1801	490000
5	4.69	37	173.53	21.9961	1369
6	4.99	50	249.5	24.9001	2500
7	3.49	500	1745	12.1801	250000
8	4.75	60	285	22.5625	3600
9	4.39	90	395.1	19.2721	8100
10	3.19	920	2934.8	10.1761	846400
Totals	**39.42**	**4,097**	**14,123**	**160.0298**	**2,661,569**

3(b)

$$\bar{x} = 3.94, \bar{y} = 409.7$$

$$\hat{\beta}_1 = \frac{xy - n\bar{x}\bar{y}}{x^2 - n\bar{x}^2}$$

$$\hat{\beta}_1 = \frac{14123 - (10)(409.7)(3.94)}{160 - (10)(3.94)^2} = \frac{-2019.2}{4.764} = -424$$

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

$$\hat{\beta}_0 = 409.7 - (-424)(3.94) = 2080$$

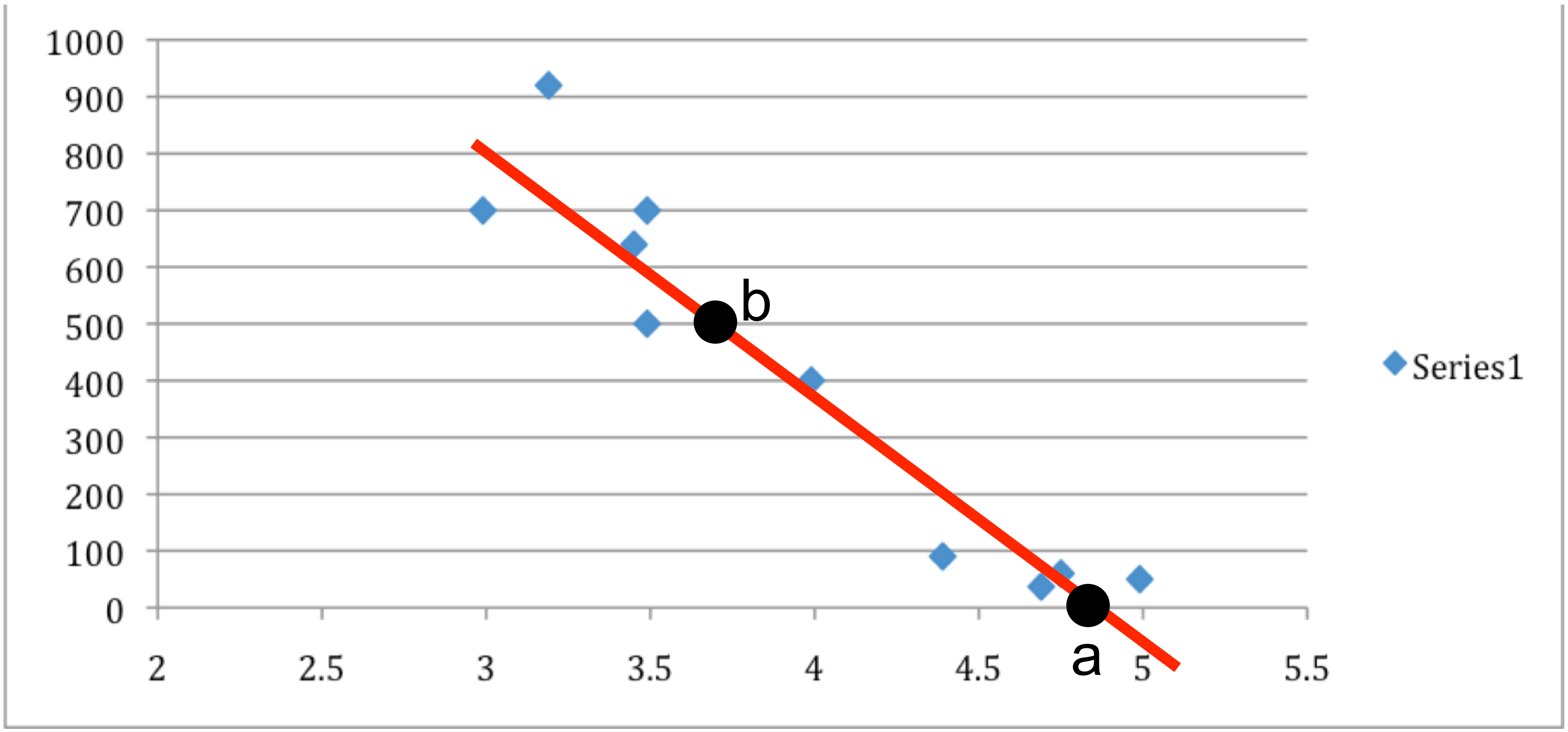
$$\hat{y} = 2080 - 424x$$

3 (b)

$$\hat{y} = 2{,}080 - 424x$$

(a) If y = 0, x = 4.9

(b) If y = 500, x = 3.7



3(c)

$$\hat{y} = 2080 - 424x$$

x = £3.50 then
Estimated y = 2080-424(3.50)=596

C2.It is ***possible*** but NOT **appropriate** to use the above equation for a price of £1.99 because the sample data for this estimated linear regression equation included prices for £2.99 to £4.99.

3(d)

$$R^2 = 1 - \frac{ESS}{TSS}$$

$$ESS = \Sigma y^2 - \hat{\beta}_0 \Sigma y - \hat{\beta}_1 xy$$

ESS = 2,661,569 – (2,080)(4,097) – (-424)(14,123) = 127,961

$$TSS = \Sigma y^2 - n\bar{y}^2$$

TSS = 2,661,569 – 10(409.7)2 = 983,028

R^2 = 1 – 127,961/983,028 = 0.87

87% of Variation in Demand (y) is Explained by Price (x).

3(e)

$$SE(\hat{\beta}_1) = \frac{S}{\sqrt{\Sigma x^2 - n\bar{x}}}$$

$$S^2 = \frac{ESS}{n-2} = \frac{127{,}961}{8} = 15{,}995$$

$$S = \sqrt{15{,}995} = 126.5$$

$$SE(\hat{\beta}_1) = \frac{126.5}{\sqrt{160 - 10(3.94)^2}} = 58$$

3(f)

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 < 0 \end{cases}$$

$$t = \frac{\hat{\beta}_1}{SE(\hat{\beta})} = \frac{-424}{58} = -7.3$$

$$t_{0.05,8} = 1.86$$

$$t = -7.3 < t_c(-1.86)$$

Thus we reject H_0.
There is a negative relationship between price & demand.

3(g)

$$C.I. = \hat{\beta}_1 \pm t_{\alpha/2} * SE(\hat{\beta}_1)$$

$$t_{\alpha/2, n-2} = t_{0.05/2, 10-2} = t_{0.025, 8} = 2.306$$

$$C.I. = -424 \pm (2.306)(58)$$

C.I.= -557 to -290

3(h)

C.I. for the mean

$$= \hat{y} \pm t_c * S\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\Sigma x^2 - n\overline{x}^2}}$$

$$= 596 \pm (2.306)(126.5)\sqrt{\frac{1}{10} + \frac{(3.50 - 3.94)^2}{160 - (10)(3.94)^2}}$$

$$= 596 \pm 109$$

= 487 to 705

3(l)

P.I.

$$= \hat{y} \pm t_c * S\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\Sigma x^2 - n\bar{x}^2}}$$

$$= 596 \pm (2.306)(126.5)\sqrt{1 + \frac{1}{10} + \frac{(3.50 - 3.94)^2}{160 - (10)(3.94)^2}}$$

$$= 596 \pm 312$$

= 284 to 908

Thank you

www.pairach.com/teaching