

Business Statistics



Level 3

Series 3 2003

(Code 3009)

Model Answers

Business Statistics Level 3

Series 3 2003

How to use this booklet

Model Answers have been developed by LCCIEB to offer additional information and guidance to Centres, teachers and candidates as they prepare for LCCIEB examinations. The contents of this booklet are divided into 3 elements:

- (1) Questions – reproduced from the printed examination paper
- (2) Model Answers – summary of the main points that the Chief Examiner expected to see in the answers to each question in the examination paper, plus a fully worked example or sample answer (where applicable)
- (3) Helpful Hints – where appropriate, additional guidance relating to individual questions or to examination technique

Teachers and candidates should find this booklet an invaluable teaching tool and an aid to success.

The London Chamber of Commerce and Industry Examinations Board provides Model Answers to help candidates gain a general understanding of the standard required. The Board accepts that candidates may offer other answers that could be equally valid.

© Education Development International plc 2003

All rights reserved; no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the Publisher. The book may not be lent, resold, hired out or otherwise disposed of by way of trade in any form of binding or cover, other than that in which it is published, without the prior consent of the Publisher.

Business Statistics Level 3

Series 3 2003

QUESTION 1

The sales figures for a company specialising in pre-packed salad have been increasing over the past 10 years.

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Sales (£000)	325	342	377	385	416	444	489	511	534	576

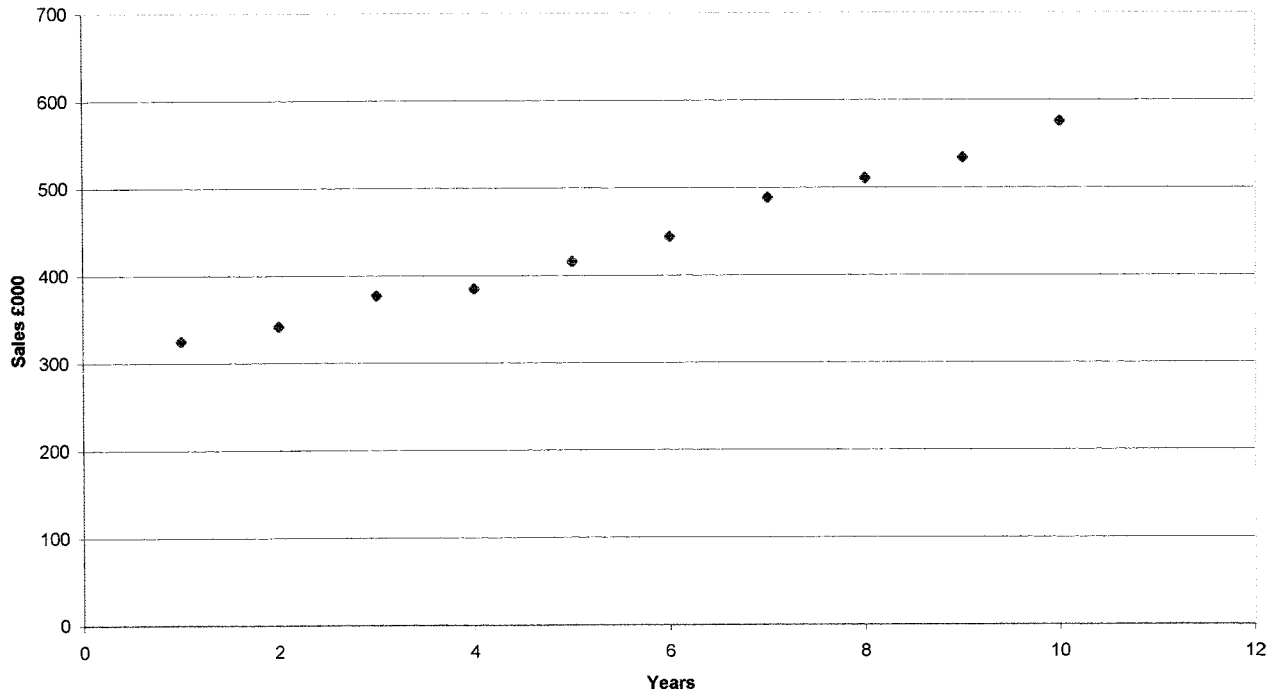
- (a) Plot the data on a scatter diagram. (5 marks)
- (b) Using the method of least squares find the regression equation. (10 marks)
- (c) Describe clearly the meaning of the coefficients of the regression equation you have found in (b) above. (3 marks)
- (d) Estimate the sales for 2004, using the equation found in (b) above. (2 marks)

(Total 20 marks)

Model Answer to Question 1

(a) $\sum x = 55, \sum x^2 = 385, \sum y = 4,399, \sum xy = 26,501$

Sales of salad



(b)
$$b = \frac{N\sum xy - \sum x \cdot \sum y}{N\sum x^2 - (\sum x)^2}$$

$$= \frac{10 \times 26,501 - 55 \times 4,399}{10 \times 385 - 55^2}$$

$$= \frac{265,010 - 241,945}{3,850 - 3,025} = \frac{23,065}{825}$$

$$= 27.9576$$

$$a = \frac{\sum y}{N} - b \frac{\sum x}{N}$$

$$= \frac{4,399}{10} - \frac{27.9576 \times 55}{10}$$

$$= 439.9 - 153.77 = 286.13$$

Sales = 286.13 + 27.96 year

(c) a (286.13) is the constant value for year 1992 (in £000)

b (27.96) is the growth rate of sales pa (in £000)

(d) 2004 = Year 12

$$\begin{aligned} \text{Estimated sales} &= 286.13 + (27.96 \times 12) \\ &= 286.13 + 335.52 \\ &= 621.65 \text{ £000} \end{aligned}$$

QUESTION 2

- (a) Explain the **advantages** and **disadvantages** of **each** of the Laspeyres and Paasche price indices.

(4 marks)

A company keeps a record of the value and volume of sales for its three key products R, S and T.

Product	1997		2002	
	Volume (000)	Value (£000)	Volume (000)	Value (£000)
R	7	26	6	29
S	5	33.7	7	34
T	12	43	13	56.8

- (b) Determine:
- (i) the Laspeyres price index
 - (ii) the Paasche price index.

Explain why the results differ.

(10 marks)

In 1992 the total value of sales on the three key products was £68,300.

- (c) Calculate an index number for the change in value of sales (with 1992 = 100) between:
- (i) 1992 and 1997
 - (ii) 1992 and 2002.

Comment on the growth in value of sales between 1992 and 1997 and 1997 and 2002.

(6 marks)

(Total 20 marks)

Model Answer to Question 2

(a) Laspeyres

Advantage. Fixed Base year quantities therefore easier to calculate.

Disadvantage. Base year quantities become outdated. May overestimate the impact of inflation.

Paasche

Current year weights therefore up-to-date but more work is required to find the new weights. May underestimate the impact of inflation.

(b) (i) and (ii)

$$\begin{aligned} 1997 \text{ value} &= \sum P_0 Q_0 = 102.7 \\ 1997 \text{ prices} &= 26/7, \quad 33.7/5 \text{ \& } 43/12 \\ &= \text{£}3.71, \quad \text{£}6.74 \text{ \& } \text{£}3.58 \end{aligned}$$

$$\begin{aligned} 2002 \text{ value} &= \sum P_1 Q_1 = 119.8 \\ 2002 \text{ prices} &= 29/6, \quad 34/7 \quad \text{\& } 56.8/13 \\ &= \text{£}4.83 \quad \text{£}4.86 \quad \text{\& } \text{£}4.37 \end{aligned}$$

$$\begin{aligned} \sum P_1 Q_0 &= 7 \times 4.83 + 5 \times 4.86 + 12 \times 4.37 \\ &= 33.81 + 24.3 + 52.44 \\ &= 110.55 \end{aligned}$$

$$\begin{aligned} \sum P_0 Q_1 &= 6 \times 3.71 + 7 \times 6.74 + 13 \times 3.58 \\ &= 22.26 + 47.18 + 46.54 \\ &= 115.98 \end{aligned}$$

$$\begin{aligned} \text{Laspeyres} &= \frac{\sum P_1 Q_0 \times 100}{\sum P_0 Q_0} = \frac{110.55 \times 100}{102.7} \\ &= 107.6 \end{aligned}$$

$$\begin{aligned} \text{Paasche} &= \frac{\sum P_1 Q_1 \times 100}{\sum P_0 Q_1} = \frac{119.8 \times 100}{115.98} \\ &= 103.3 \end{aligned}$$

Comment – because the weights differ

(c) (i) and (ii)

$$\begin{array}{llll} 1992 & 68,300 & & = 100 \\ 1997 & 102,700 & 102,700/68,300 \times 100 & = 150.4 \\ 2002 & 119,800 & 119,800/68,300 \times 100 & = 175.4 \end{array}$$

Comment – the rate of increase has reduced between the first five year period and the second five year period.

QUESTION 3

A company has 2 sales regions, North and South. A random sample was taken of sales personnel in each area and their sales figures for the month of June were recorded.

North (£000)	48	53	95	65	73	81	46	
South (£000)	87	75	63	58	68	72	84	55

- (a) Test whether there is evidence to support the claim that the average sales in the 2 regions are different.

(14 marks)

- (b) Explain the difference between a **Type I** and a **Type II** error. Which type of error could have been committed in (a) above?

(6 marks)

(Total 20 marks)

Model Answer to Question 3

(a) Null hypothesis: there is no difference in the sales between the 2 regions.

Alternative hypothesis: there is a difference in the sales between the 2 regions.

Degrees of Freedom $n_1 + n_2 - 2 = 7 + 8 - 2 = 13$

Critical t 0.05 = 2.16 .t0.01. $x = 3.01$

$$\sum X_N = 461 \qquad \sum X_S = 562$$

$$\bar{x}_N = 65.857 \qquad \bar{x}_S = 70.25$$

$$\sum X_N^2 = 32369 \qquad \sum X_S^2 = 40,416$$

$$\text{or } \sum (X - \bar{x}_N)^2 = 2008.86 \quad \sum (X - \bar{x}_S)^2 = 935.5$$

$$SD_N = 30 \qquad SD_S = 11.56$$

$$\text{Pooled SD} = \sqrt{\frac{\sum(x - \bar{x}_N)^2 + \sum(x - \bar{x}_S)^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{2,008.86 + 935.5}{7 + 8 - 2}}$$

$$= \sqrt{\frac{2,944.36}{13}} = 15.05$$

$$t = \frac{\bar{x}_N - \bar{x}_S}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{65.857 - 70.25}{\sqrt{\frac{1}{8} + \frac{1}{7}} \times 15.05} = \frac{-4.393}{7.789}$$

$$= -0.56$$

Conclusion: The calculated value of t is less than the critical value of t therefore accept the null hypothesis. The sales in the 2 regions do not differ.

(b) A Type I error is the error of rejecting the null hypothesis when it is true.

A Type II error is the error accepting the null hypothesis when it is false.

A Type II error may have occurred.

QUESTION 4

- (a) The probabilities that a company is successful when it bids for its next 3 contracts, A, B and C, are 0.3, 0.25 and 0.1 respectively.

Calculate the probability that:

- (i) no bid is successful
- (ii) one bid is successful
- (iii) given one bid is successful, it is for Contract B.

(10 marks)

- (b) (i) Why might the rank correlation coefficient be used in preference to the product-moment correlation coefficient?

(4 marks)

- (ii) The price of petrol is recorded at 10 garages which are also appraised for quality of service. The data are given below:

Garage	A	B	C	D	E	F	G	H	I	J
Petrol price	73	66	68	74	69	73	68	79	70	67
Quality rank	3	5	6	2	8	1	4	7	9	10

Assume that 1 is the best quality

Calculate the rank correlation coefficient between the petrol price and quality rank. Interpret your answer.

(6 marks)

(Total 20 marks)

Model Answer to Question 4

(a) (i) Not A = $1 - 0.03 = 0.7$

Not B = $1 - 0.25 = 0.75$

Not C = $1 - 0.1 = 0.9$

No bid successful = $0.7 \times 0.75 \times 0.9 = 0.4725$

(ii) One successful bid $0.7 \times 0.75 \times 0.1 = 0.0525$

+ $0.7 \times 0.25 \times 0.9 = 0.1575$

+ $0.3 \times 0.75 \times 0.9 = \underline{0.2025}$

0.4125

(iii) $\frac{\text{Probability bid B successful}}{\text{Probability/success}} = \frac{0.1575}{0.4125}$

= 0.3818

(b) (i) Rank correlation is easier and quicker to calculate than product-moment correlation. One or both of the sets of data may be qualitative and therefore only rankable. Only secondary ranked data may be available.

(ii) Garage	A	B	C	D	E	F	G	H	I	J
Petrol Rank	3½	10	7½	2	6	3½	7½	1	5	9
Quality Rank	3	5	6	2	8	1	4	7	9	10
Difference d	½	5	1½	0	2	2½	3½	6	4	1
Difference ² d ²	¼	25	2¼	0	4	6¼	12¼	36	16	1

$\Sigma d^2 = 103$

$$R = 1 - \frac{6 \Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 103}{10(100 - 1)}$$

= $1 - 0.624 = 0.376$

Comment. There is weak positive association between petrol price and quality.

QUESTION 5

(a) Give 2 examples of the use of the normal distribution in a business context.

(2 marks)

The average period for receipt of payment after invoices have been sent out is 45 days with a standard deviation of 10 days. You may assume that the distribution of payment time is normally distributed and continuous.

(b) Find the probability that invoices are paid:

- (i) within 30 days of the invoices being sent out
- (ii) between 30 and 50 days after the invoices are sent out
- (iii) more than 50 days after the invoices are sent out.

(8 marks)

The average invoice amount is £5,500 and 2,000 invoices are sent out each month.

If payment is received within 30 days, a discount of 5% is allowed on the account.

If payment is received between 30 and 50 days, a discount of 2½% is allowed.

If payment is received after more than 50 days, the full amount is payable.

(c) How much, on average, can the company expect to receive from a month's invoices?

(7 marks)

(d) Within how many days will 90% of the invoices be paid?

(3 marks)

(Total 20 marks)

Model Answer to Question 5

(a) Setting working and action limits for quality control charts.

Estimation of probabilities eg sales of a magazine or a newspaper.

$$(b) (i) Z = \frac{x - \mu}{\sigma} = \frac{30 - 45}{10} = \frac{-15}{10} = -1.5$$

$$\begin{aligned} \text{prob} (< 30) &= 1 - 0.933 \\ &= 0.067 \end{aligned}$$

$$(ii) Z = \frac{x - \mu}{\sigma} = \frac{50 - 45}{10} = 0.5$$

$$p(45 < 50) = 0.692 - 0.5 = 0.192$$

$$p(30 < 45) = 0.933 - 0.5 = 0.433$$

$$p(30 < 50) = 0.625$$

$$(iii) p(> 50) = 1 - 0.692 = 0.308$$

$$(c) \text{ Less than 30 days} = 0.067 \times 5,500 \times 0.95 \times 2,000 = 700,159$$

$$30 < 50 \text{ days} = 0.625 \times 5,500 \times 0.975 \times 2,000 = 6,703,125$$

$$> 50 \text{ days} = 0.308 \times 5,500 \times 1 \times 2,000 = 3,388,080$$

$$\begin{aligned} &£10,791,275 \div 2,000 \\ &£5,395.64 \end{aligned}$$

(d) From tables Z value of 1.3 gives 90.3%

$$+ 1.3 = \frac{x - 45}{10} = 13 + 45 = x = 58 \text{ days}$$

QUESTION 6

During 2002 a random sample of employees was investigated to ascertain the amount of absenteeism that exists within a company. The data collected are summarised in the following table:

Days absent	Number of employees
0	21
1 to 4	34
5 to 10	53
11 to 15	46
16 to 25	29
26 to 40	17

- (a) Calculate the arithmetic mean and the standard deviation of the number of days absent. (8 marks)
- (b) (i) Explain what is meant by the 99% confidence interval.
- (ii) Calculate a 99% confidence interval for the mean number of days absent in 2002. (6 marks)

In 2001 the mean number of days absent was 9.13.

- (c) Test whether there has been a significant increase in the mean number of days absent between 2001 and 2002. (6 marks)

(Total 20 marks)

Model Answer to Question 6

(a)	mpt(x)	f	fx	fx ²
0	0	21	0	0
1 to 4	2.5	34	85	212.5
5 to 10	7.5	53	397.5	2,981.25
11 to 15	13	46	598	7,774
16 to 25	20.5	29	594.5	12,187.25
26 to 40	33	<u>17</u>	<u>561</u>	<u>18,513</u>
		<u>200</u>	<u>2,236</u>	<u>41,668</u>
		Σf	Σfx	Σfx^2

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{2,236}{200} = 11.18 \text{ days}$$

$$s = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{41,668}{200} - 11.18^2} = 9.1295 \text{ (9.13) days}$$

(b) (i) The confidence interval provides an interval within which you would expect the true population mean to lie with 95% confidence.

(ii) 99% ci $\mu = \bar{x} \pm 2.58 \frac{s}{\sqrt{n}}$

$$= \mu = 11.18 \pm 2.58 \times \frac{9.13}{\sqrt{200}}$$

$$= 9.514 \text{ to } 12.846 \text{ (9.501 to 12.859) if 2.6 used}$$

(c) Null hypothesis: there is no difference in the mean number of days absent.

Alternative hypothesis: the mean number of days absent has increased.

Critical z = 1.64/2.33

$$Z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{11.18 - 9.13}{9.13/\sqrt{200}}$$

$$= \frac{2.05}{0.646} = 3.17$$

Conclusion. Reject the null hypothesis. Accept the alternative hypothesis, there has been an increase in the mean number of days absent.

QUESTION 7

- (a) A random sample of 590 computer users was taken to identify the main use of the machines. The results are shown in the table below:

Age of User	Usage		
	Internet	Games	Email
Under 30	120	80	100
30 to 50	60	40	80
Over 50	40	30	40

Test whether the evidence supports the view that age and type of computer use are related.

(12 marks)

- (b) In a previous survey, the national percentages for main computer usage were:

Usage		
Internet	Games	Email
35%	25%	40%

By combining the age groups, test whether the pattern of computer use has changed between the 2 surveys.

(8 marks)

(Total 20 marks)

Model Answer to Question 7

(a) Null hypothesis: no association between age and computer usage.

Alternative hypothesis: there is association between age and computer usage.

Degrees of Freedom = $(R - 1) \times (C - 1) = (3 - 1) \times (3 - 1) = 4$

Critical χ^2 9.49/13.28

Expected	111.86	76.27	111.86
Values	67.12	45.76	67.12
	41.02	27.97	41.02

Contribution 0.592 + 0.182 + 1.258 +

to χ^2 0.755 + 0.726 + 2.472 +

0.025 + 0.148 + 0.025

$\chi^2 = 6.184$

Calculated χ^2 6.184 < 9.49, therefore there is no association between age and computer usage.

(b) Null hypothesis: computer use has not changed between the surveys.

Alternative hypothesis: computer use has changed between the surveys.

Degrees of freedom = $N - 1 = 3 - 1 = 2$

Critical $\chi^2 = 5.99/9.21$

Expected value	206.5	147.5	236
Contributions to χ^2	0.893	0.042	1.085

$\chi^2 = 2.01$

Conclusions. Calculated χ^2 is less than critical χ^2

Accept the null hypothesis. Computer use has not changed between the surveys.

QUESTION 8

(a) Explain the difference between a **census** and a **sample**. (2 marks)

(b) Explain what is meant by the **standard error of the proportion**. (4 marks)

A company is concerned about the difference in the performance of the Northern and Southern sales teams. A random sample of 70 sales people in the North showed 55 achieving their sales target, whilst in the South a random sample of 90 sales people showed 73 achieving their sales target.

(c) Test whether the proportion of sales people achieving their sales target is greater in the South than in the North. (14 marks)

(Total 20 marks)

Model Answer to Question 8

(a) A census surveys all of the population.

Sample surveys a proportion of the population.

(b) The standard error of the proportion is the standard deviation of the proportions estimated from samples of a given size. It is equal to $\sqrt{pq/n}$

(c) Null hypothesis: there is no difference in the proportion achieving their sales target.

Alternative hypothesis: there is a difference in the proportion achieving their sales target.

Critical $z = 1.96$

P north = $55/70 = 0.7857$

P south = $73/90 = 0.8111$

Pooled $p = \frac{55 + 73}{70 + 90} = \frac{128}{160} = 0.8$

$$z = \frac{P_n - P_s}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
$$= \frac{0.7857 - 0.8111}{\sqrt{0.8 \times 0.2 (0.01429 + 0.01111)}}$$
$$= \frac{-0.0254}{0.06375} = -0.398 \quad (-0.4)$$

Critical $z = 1.96$

Conclusion: Accept the null hypothesis. There is no difference in the proportion achieving their sales target.

Mean $\bar{x} = \frac{\sum fx}{\sum f}$

Standard deviation $s = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$

Pearson measure of skewness $\frac{3(\bar{x} - \text{Median})}{s}$

Coefficient of variation $\frac{s}{\bar{x}} \times 100$

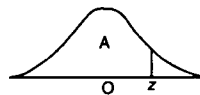
Product moment correlation coefficient $r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{(n\sum x^2 - (\sum x)^2)(n\sum y^2 - (\sum y)^2)}}$

Spearman's rank correlation coefficient $r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$

Least squares regression line $\hat{y} = a + bx$
 $b = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$
 $a = \frac{\sum y}{n} - \frac{b\sum x}{n}$

TABLE 1 – NORMAL DISTRIBUTION

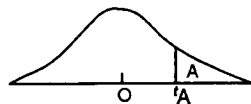
A is the area to the left of the given value of z



z	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
A	.500	.540	.580	.618	.655	.692	.726	.758	.788	.816
z	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
A	.841	.864	.885	.903	.919	.933	.945	.955	.964	.971
z	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9
A	.977	.982	.986	.989	.992	.994	.995	.996	.997	.998

TABLE 2 – t DISTRIBUTION

t_A is the value of the t statistic with ν degrees of freedom with area A to the right of it



ν	1	2	3	4	5	6	7	8
t _{0.05}	6.31	2.92	2.35	2.13	2.02	1.94	1.90	1.86
t _{0.025}	12.71	4.30	3.18	2.78	2.57	2.45	2.37	2.31
t _{0.01}	31.82	6.97	4.54	3.75	3.37	3.14	3.00	2.90
t _{0.005}	63.66	9.93	5.84	4.60	4.03	3.71	3.50	3.36
ν	9	10	11	12	13	14	15	16
t _{0.05}	1.83	1.81	1.80	1.78	1.77	1.76	1.75	1.75
t _{0.025}	2.26	2.23	2.20	2.18	2.16	2.15	2.13	2.12
t _{0.01}	2.82	2.76	2.72	2.68	2.65	2.62	2.60	2.58
t _{0.005}	3.25	3.17	3.11	3.05	3.01	2.98	2.95	2.92
ν	17	18	19	20	21	22	23	24
t _{0.05}	1.74	1.73	1.73	1.72	1.72	1.71	1.71	1.71
t _{0.025}	2.11	2.10	2.09	2.09	2.08	2.07	2.07	2.06
t _{0.01}	2.57	2.55	2.54	2.53	2.52	2.51	2.50	2.49
t _{0.005}	2.90	2.88	2.86	2.85	2.83	2.82	2.81	2.80

One sample z test

Mean $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ Proportion $z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}}$

Two sample z test

Mean $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ Proportion $z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

where $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$

Price Quantity

Laspeyres index $\frac{\sum P_1 Q_0}{\sum P_0 Q_0} \times 100$ $\frac{\sum P_0 Q_1}{\sum P_0 Q_0} \times 100$

Paasche index $\frac{\sum P_1 Q_1}{\sum P_0 Q_1} \times 100$ $\frac{\sum P_1 Q_1}{\sum P_1 Q_0} \times 100$

Weighted index $\frac{\sum WI}{\sum W}$

One sample t test

$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ where $s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$

Independent samples t test

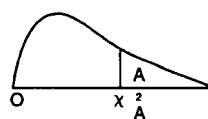
$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n} + \frac{1}{m}}}$ where $s = \sqrt{\frac{\sum(x - \bar{x})^2 + \sum(y - \bar{y})^2}{n + m - 2}}$

Chi-square test $\chi^2 = \sum \frac{(O - E)^2}{E}$

Test for $\rho = 0$ $t = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}}$

TABLE 3 – CHI SQUARED DISTRIBUTION

χ²_A is the value of the χ² statistic with ν degrees of freedom with area A to the right of it



ν	1	2	3	4	5	6
χ ² _{0.05}	3.84	5.99	7.81	9.49	11.07	12.59
χ ² _{0.01}	6.63	9.21	11.34	13.28	15.09	16.81
ν	7	8	9	10	11	12
χ ² _{0.05}	14.07	15.51	16.92	18.31	19.68	21.03
χ ² _{0.01}	18.48	20.09	21.67	23.21	24.73	26.22



**EXAMINATIONS
BOARD**

London Chamber of Commerce and Industry Examinations Board

Athena House 112 Station Road

Sidcup Kent DA15 7BJ United Kingdom

Facsimile: +44 (0) 20 8302 4169/+44 (0) 20 8309 5169

Website: www.lccieb.com

Customer Service Team Tel: +44 (0) 20 8309 3000 email: custserv@lccieb.org.uk

© Education Development International plc 2003