University of Kelaniya – Sri Lanka

Centre for Distance and Continuing Education

Faculty of Commerce and Management Studies

Bachelor of Business Management (General) Second Year Examination (External) – 2022

December - 2024

BMGT E2045 – Statistics for Management

No. of Questions: Eight (08) Time: 03 hours

Answer any five (05) Questions.

Non Programmable Calculators are allowed.

Question No. 01

a) Define statistics and write down five major characteristics of statistics.

(08 Marks)

b) Explain three primary data collection methods.

(06 Marks)

c) Describe any three uses of statistics.

(06 Marks)

(Total 20 Marks)

Question No. 02

a) Differentiate between descriptive statistics and inferential statistics, providing appropriate examples.

(08 Marks)

b) Explain discrete and continuous variables, providing an example for each.

(06 Marks)

c) State whether each variable given below is qualitative or quantitative.

i. Height in inches
ii. Gender
iii. Race
iv. Test scores
v. Language
vi. Age in years

(06 Marks)

(Total 20 Marks)

Question No. 03

a) The following grouped frequency distribution gives the monthly wages of 200 employees in a computer firm. Calculate the mean value and interpret your result.

Wages (in '000 rupees)	Number of employees
500 and less than 550	4
550 and less than 600	26
600 and less than 650	133
650 and less than 700	35
700 and less than 750	2

(08 marks)

b) i. What is dispersion?

(04 marks)

ii. Name four measures of dispersion.

(04 marks)

iii. List the main characteristics of each measure.

(04 marks)

(Total 20 Marks)

Question No. 04

Many internet users shopped online during the holiday season. In a survey of 1000 customers who did hodilay shopping online, 34% indicated that they were not satisfied with their experience. Of the customers that were not satisfied, 24% indicated that they did not receive their product in time for the hodilays. Suppose the following complete set of results was reported:

Satisfied with Experience	**Received products in time for holidays**		
	Yes	No	Total
Yes	560	100	660
No	260	80	340
Total	820	180	1000

Using above information, answer the following questions.

a) Give an example of a simple event.

(02 marks)

b) Give an example of a joint event.

(02 marks)

c) What is the complement of " satisfied with experience"?

(02 marks)

d) If a customer is selected at random , what is the probability that he/ she,

i. is satisfied with experience?

(02 marks)

ii. received the product in time for the holidays?

(02 marks)

iii. is satisfied with the experience and did not receive the product in time for the holidays?

(02 marks)

iv. is not satisfied with the experience or did not receive the product in time for the holidays?

(02 marks)

e) Suppose the customer chosen did not receive the product in time for the holidays, then what is the probability that she/he satisfied with experience.

(02 marks)

f) Suppose the customer chosen is satisfied with experience, then what is the probability that she/he received the product in time for the holidays?

(02 marks)

g) Are satisfied with experience and received the product in time for the holidays statistically independent or not? Explain.

(02 marks)

(Total 20 Marks)

Question No. 05

a) What are the properties of normal disrtibutions?

(05 marks)

b) Find the value of z for a standard normal distribution, such that the area in the left tail is 0.05.

(05 marks)

c) Suppose IQ score are normally distributed with a mean of 100 and a standard deviation of 15. Then,

i. What percentage of people has an IQ score above 120?

(05 marks)

ii. What is the maximum IQ score for the bottom 30%?

(05 marks)

(Total 20 Marks)

Question No. 06

a) Suppose an accounting firm conducts a study to determine the time required to complete an individual's tax forms. It randomly surveys 100 people, finding a sample mean of 23.6 hours. The standard deviation is known to be 7.0 hours, and the population distribution is assumed to be normal.

i. Construct a 95% confidence interval based on the given information.

(05 marks)

ii. Interpret your answer.

(05 marks)

b) To estimate the proportion of female students at a university, a random sample of 120 students is selected, and 69 of them are female.

i. Construct a 90% confidence interval for the above information.

(05 marks)

ii. Interpret your answer.

(05 marks)

(Total 20 Marks)

Question No. 07

a) What is the key difference between simple regression and multiple regression analysis? Explain your answer by providing examples.

(05 marks)

b) A researcher is interested in finding out how price and promotional expenditure affect the sales of chocolate bars sold by company XYZ. Price, promotional expenditure, and sales are measured in rupees. Data were collected from a sample of 50 customers over a period of one month. You are provided with the following regression (Excel) output. You are required to answer the questions based on this output.

Regression Statistics

Multiple R	0.959523
R Square	0.920684
Adjusted R Square	0.917884
Standard Error	4.397575
Observations	50

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	3204.229	1602.115	82.84515	.0000134
Residual	47	135.3707	19.33867		
Total	49	3339.6			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	23.37009	50.63295	0.876309	0.409924	-75.3578	164.098
Promotional Expenditure (Rs)	72.14165	2.569993	2.807076	0.026256	1.137097	13.29123
Price (Rs per unit)	-10.10404	0.175523	-0.59274	0.057199	-0.51908	0.311006

i. Identify the independent and dependent variables.

(03 marks)

ii. Write down and interpret the regression equation.

(03 marks)

iii. What does the significance of F value mean?

(03 marks)

iv. What does the model predict for sales of chocolate bars with Rs 1000 of expenditure in promotions and Rs. 500 of the price?

(03 marks)

v. What is the interpretation of the R^2?

(03 marks)

(Total 20 Marks)

Question No. 08

a) Based on the available data, it is observed that 400 out of 850 customers purchased groceries online. Can we conclude that most customers (more than half of customers) are shifting towards online shopping for groceries? (Hint: Develop hypotheses, test hypotheses, make a decision, interpret the conclusion.)

(10 marks)

b) A telecom service provider claims that customers spend an average of Rs.400 per month, with a standard deviation of Rs. 25. However, a random sample of 50 customer bills shows a mean of Rs.250 and a standard deviation of Rs.15. Does this sample data support the service provider's claim? ? (Hint: develop hypotheses, test hypotheses, make decision, interpret conclusion)

(10 marks)

(Total 20 Marks)

FORMULAE

Summary Measures

Sample Mean

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

Sample Standard Deviation

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n-1}}$$

Probability Rules

- **Complement rule**

  $P(A^C) = 1 - P(A)$

- **Addition rule**

  General: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

  For independent events:

  $P(A \text{ or } B) = P(A) + P(B) - P(A)P(B)$

  For mutually exclusive events: $P(A \text{ or } B) = P(A) + P(B)$

- **Multiplication rule**

  General: $P(A \text{ and } B) = P(A)P(B \mid A)$

  For independent events: $P(A \text{ and } B) = P(A)P(B)$

  For mutually exclusive events: $P(A \text{ and } B) = 0$

- **Conditional Probability**

  General: $P(A \mid B) = \dfrac{P(A \text{ and } B)}{P(B)}$

  For independent events: $P(A \mid B) = P(A)$

  For mutually exclusive events: $P(A \mid B) = 0$

Discrete Random Variables

Mean

$$E(X) = \mu = \sum x_i p_i = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k$$

Standard Deviation

$$s.d.(X) = \sigma = \sqrt{\sum (x_i - \mu)^2 p_i} = \sqrt{\sum \left(x_i^2 p_i\right) - \mu^2}$$

Binomial Random Variables

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where $\dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$

Mean

$$E(X) = \mu_X = np$$

Standard Deviation

$$s.d.(X) = \sigma_X = \sqrt{np(1-p)}$$

Normal Random Variables

- $z-\text{score} = \dfrac{\text{observation} - \text{mean}}{\text{standard deviation}} = \dfrac{x - \mu}{\sigma}$
- Percentile: $x = z\sigma + \mu$
- If X has the $N(\mu, \sigma)$ distribution, then the variable

  $Z = \dfrac{X - \mu}{\sigma}$ has the $N(0,1)$ distribution.

Normal Approximation to the Binomial Distribution

If X has the $B(n, p)$ distribution and the sample size n is large enough (namely $np \geq 10$ and $n(1-p) \geq 10$),

then X is approximately $N\left(np, \sqrt{np(1-p)}\right)$.

Sample Proportions

$$\hat{p} = \frac{x}{n}$$

Mean

$$E(\hat{p}) = \mu_{\hat{p}} = p$$

Standard Deviation

$$\text{s.d.}(\hat{p}) = \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Sampling Distribution of $\hat{p}$

If the sample size n is large enough (namely, $np \geq 10$ and $n(1-p) \geq 10$)

then $\hat{p}$ is *approximately* $N\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right)$.

Sample Means

Mean

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

Standard Deviation

$$s.d.(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Sampling Distribution of $\bar{X}$

If X has the $N(\mu, \sigma)$ distribution, then $\bar{X}$ is

$$N(\mu_{\bar{X}}, \sigma_{\bar{X}}) \Leftrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

If X follows *any* distribution with mean μ and standard deviation σ and n is large,

then $\bar{X}$ is *approximately* $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

This last result is **Central Limit Theorem**

Population Proportion

Parameter p

Statistic $\hat{p}$

Standard Error

$$\text{s.e.}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence Interval

$$\hat{p} \pm z^* \text{s.e.}(\hat{p})$$

Conservative Confidence Interval

$$\hat{p} \pm \frac{z^*}{2\sqrt{n}}$$

Large-Sample *z*-Test

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Sample Size

$$n = \left(\frac{z^*}{2m}\right)^2$$

Two Population Proportions

Parameter $p_1 - p_2$

Statistic $\hat{p}_1 - \hat{p}_2$

Standard Error

$$\text{s.e.}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Confidence Interval

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \text{s.e.}(\hat{p}_1 - \hat{p}_2)$$

Large-Sample *z*-Test

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

Population Mean

Parameter μ

Statistic $\bar{x}$

Standard Error

$$\text{s.e.}(\bar{x}) = \frac{s}{\sqrt{n}}$$

Confidence Interval

$$\bar{x} \pm t^* \text{s.e.}(\bar{x}) \qquad \text{df} = n-1$$

Paired Confidence Interval

$$\bar{d} \pm t^* \text{s.e.}(\bar{d}) \qquad \text{df} = n-1$$

One-Sample *t*-Test

$$t = \frac{\bar{x} - \mu_0}{\text{s.e.}(\bar{x})} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \qquad \text{df} = n-1$$

Paired *t*-Test

$$t = \frac{\bar{d} - 0}{\text{s.e.}(\bar{d})} = \frac{\bar{d}}{s_d/\sqrt{n}} \qquad \text{df} = n-1$$

Two Population Means

General

Parameter $\mu_1 - \mu_2$

Statistic $\bar{x}_1 - \bar{x}_2$

Standard Error

$$\text{s.e.}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Confidence Interval

$$(\bar{x}_1 - \bar{x}_2) \pm t^*\left(\text{s.e.}(\bar{x}_1 - \bar{x}_2)\right) \qquad \text{df} = \min(n_1 - 1, n_2 - 1)$$

Two-Sample *t*-Test

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\text{s.e.}(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \text{df} = \min(n_1 - 1, n_2 - 1)$$

Pooled

Parameter $\mu_1 - \mu_2$

Statistic $\bar{x}_1 - \bar{x}_2$

Standard Error

$$\text{pooled s.e.}(\bar{x}_1 - \bar{x}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

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

Confidence Interval

$$(\bar{x}_1 - \bar{x}_2) \pm t^*\left(\text{pooled s.e.}(\bar{x}_1 - \bar{x}_2)\right) \qquad \text{df} = n_1 + n_2 - 2$$

Pooled Two-Sample *t*-Test

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\text{pooled s.e.}(\bar{x}_1 - \bar{x}_2)} = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \text{df} = n_1 + n_2 - 2$$

One-Way ANOVA

$$\text{SS Groups} = \text{SSG} = \sum_{\text{groups}} n_i(\bar{x}_i - \bar{x})^2$$

$$\text{MS Groups} = \text{MSG} = \frac{\text{SSG}}{k-1}$$

$$\text{SS Error} = \text{SSE} = \sum_{\text{groups}} (n_i - 1)s_i^2$$

$$\text{MS Error} = \text{MSE} = s_p^2 = \frac{\text{SSE}}{N-k}$$

$$\text{SS Total} = \text{SSTO} = \sum_{\text{values}} (x_{ij} - \bar{x})^2$$

$$F = \frac{\text{MS Groups}}{\text{MS Error}}$$

ANOVA Table

Source	SS	DF	MS	F
Groups	SS Groups	$k-1$	MS Groups	F
Error	SS Error	$N-k$	MS Error	
Total	SSTO	$N-1$		

Confidence Interval $\bar{x}_i \pm t^* \frac{s_p}{\sqrt{n_i}}$ $\quad \text{df} = N-k$

Under H_0, the F statistic follows an $F(k-1, N-k)$ distribution.

Regression

Linear Regression Model

Population Version:

Mean: $\mu_Y(x) = E(Y) = \beta_0 + \beta_1 x$

Individual: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

where ε_i is $N(0, \sigma)$

Sample Version:

Mean: $\hat{y} = b_0 + b_1 x$

Individual: $y_i = b_0 + b_1 x_i + e_i$

Parameter Estimators

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{\sum(x-\bar{x})y}{\sum(x-\bar{x})^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

Residuals

$e = y - \hat{y} =$ observed y – predicted y

Correlation and its square

$$r = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

$$r^2 = \frac{SSTO - SSE}{SSTO} = \frac{SSREG}{SSTO}$$

where $SSTO = S_{YY} = \sum(y-\bar{y})^2$

Estimate of σ

$$s = \sqrt{MSE} = \sqrt{\frac{SSE}{n-2}} \quad \text{where } SSE = \sum(y-\hat{y})^2 = \sum e^2$$

Standard Error of the Sample Slope

$$\text{s.e.}(b_1) = \frac{s}{\sqrt{S_{XX}}} = \frac{s}{\sqrt{\sum(x-\bar{x})^2}}$$

Confidence Interval for β_1

$b_1 \pm t^* \text{s.e.}(b_1)$ df $= n-2$

t-Test for β_1

To test $H_0 : \beta_1 = 0$

$t = \frac{b_1 - 0}{\text{s.e.}(b_1)}$ df $= n-2$

or $F = \frac{MSREG}{MSE}$ df $= 1, n-2$

Confidence Interval for the Mean Response

$\hat{y} \pm t^* \text{s.e.(fit)}$ df $= n-2$

where $\text{s.e.(fit)} = s\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{XX}}}$

Prediction Interval for an Individual Response

$\hat{y} \pm t^* \text{s.e.(pred)}$ df $= n-2$

where $\text{s.e.(pred)} = \sqrt{s^2 + (\text{s.e.(fit)})^2}$

Standard Error of the Sample Intercept

$$\text{s.e.}(b_0) = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}}$$

Confidence Interval for β_0

$b_0 \pm t^* \text{s.e.}(b_0)$ df $= n-2$

t-Test for β_0

To test $H_0 : \beta_0 = 0$

$t = \frac{b_0 - 0}{\text{s.e.}(b_0)}$ df $= n-2$

Chi-Square Tests

Test of Independence & Test of Homogeneity

Expected Count

$$E = \text{expected} = \frac{\text{row total} \times \text{column total}}{\text{total } n}$$

Test Statistic

$$X^2 = \sum\frac{(O-E)^2}{E} = \sum\frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

df $= (r-1)(c-1)$

Test for Goodness of Fit

Expected Count

$E_i = \text{expected} = np_{i0}$

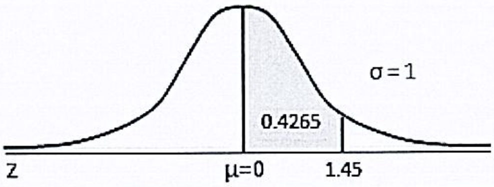
Test Statistic

$$X^2 = \sum\frac{(O-E)^2}{E} = \sum\frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

df $= k-1$

If Y follows a $\chi^2(df)$ distribution, then $E(Y) = df$ and $Var(Y) = 2(df)$.

This table provides the area between the mean and some Z score.
For example, when Z score = 1.45 the area = 0.4265.



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000