PGDM 2014-16
Research Methodology (DM 206)

Trimester–II, End-Term Examination: December 2014

Time allowed: 2 hrs 30 min **Max Marks: 50**

Roll No. ________

Instruction: Students are required to write their Roll No on every page of the question paper; writing anything except the Roll No will be treated as resorting to **Unfair Means**. For rough work, please use your answer booklet.

All workings / derivations should be clearly shown in your answers.

Required Tables provided with this Question booklet (Page 7)

Section A

Attempt ANY 3 from the 5 questions in this section. **(3 x 5)**
Each question carries 5 marks.

A1. Briefly describe the inter-related steps related to conducting research. [5]

A2. In a study of automobile traffic and air pollution, air samples taken at four different times and five different locations were analyzed to obtain the amount of particulate matter present in the air. The sample variances of particulate matter were 19.86 and 22.07 respectively for *time* and *location*. The overall sample standard deviation of particulate matter, however was found to be only 13.71.

Is there any significant difference in true average amount of particulate matter present in the air due to either different sampling times or to different locations? Use $\alpha = .05$. [5]

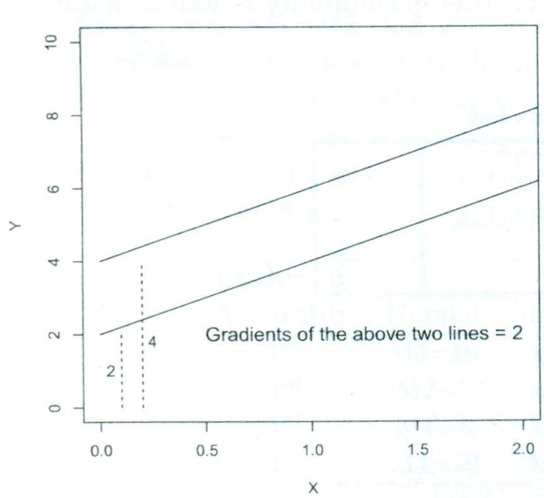
A3. Write short notes on

(a) validity and reliability [3]

(b) random error and bias [2]

A4. Explain the essential characteristics of the scientific method. [5]

A5. Write and describe the linear model depicted in the diagram below. [5]



Section B

Attempt ANY 2 from the 3 questions in this section. (2 x 10)
Each question carries 10 marks.

B1. An aluminum master alloy manufacturer produces grain refiners in ingot form. The company produces the product in four furnaces. Each furnace is known to have its own unique operating characteristics, so any experiment run in the foundry that involves more than one furnace will consider furnaces as a nuisance variable. The process engineers suspect that stirring rate affects the grain size of the product. Each furnace can be run at four different stirring rates. An experimental design is run for a particular refiner, and the resulting grain size data is as follows.

	Furnace			
Stirring rate (rpm)	1	2	3	4
5	8	4	5	6
10	14	5	6	9
15	14	6	9	2
20	17	9	3	6

Table 1: Experimental Data for Question B1.

(a). What is the treatment in this experiment and what are the levels?

(b). What type of experimental design is being used here?

(c). What is the null hypothesis in this experiment?

(d). Is there strong enough evidence at a significance of 0.05 that the *stirring rate* affects grain size?

B2. A factor analysis was done on 1428 feedback ratings by students for faculty on 12 feedback questions – variables (attributes) – viz. item13, item14, ······ item24.

Referring to the outputs from the SPSS factor analysis run with reference to the above study, write explanatory notes on any three of the following aspects.

(a) dimensionality reduction, communalities and extracted factors

(b) extraction and rotated sum of squared loadings

(c) factor analysis as distinct from principal component analysis

(d) suitability of data for Factor Analysis - overall suitability as well as suitability of individual variables

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.[a]		.934
Bartlett's Test of Sphericity[b]	Approx. Chi-Square	8676.712
	df	66
	Sig.	.000

Total Variance Explained

Factor[a]	Initial Eigenvalues[b]			Extraction Sums of Squared Loadings[f]			Rotation Sums of Squared Loadings[g]		
	Total[c]	% of Variance[d]	Cumulative %[e]	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.249	52.076	52.076	5.851	48.759	48.759	2.950	24.583	24.583
2	1.229	10.246	62.322	.806	6.719	55.478	2.655	22.127	46.710
3	.719	5.992	68.313	.360	3.000	58.478	1.412	11.769	58.478
4	.613	5.109	73.423						
5	.561	4.676	78.099						
6	.503	4.192	82.291						
7	.471	3.927	86.218						
8	.389	3.240	89.458						
9	.368	3.066	92.524						
10	.328	2.735	95.259						
11	.317	2.645	97.904						
12	.252	2.096	100.000						

Extraction Method: Principal Axis Factoring.

Communalities[a]

	Initial[b]	Extraction[c]
item13 INSTRUC WELL PREPARED	.564	.676
item14 INSTRUC SCHOLARLY GRASP	.551	.619
item15 INSTRUCTOR CONFIDENCE	.538	.592
item16 INSTRUCTOR FOCUS LECTURES	.447	.468
item17 INSTRUCTOR USES CLEAR RELEVANT EXAMPLES	.585	.623
item18 INSTRUCTOR SENSITIVE TO STUDENTS	.572	.679
item19 INSTRUCTOR ALLOWS ME TO ASK QUESTIONS	.456	.576
item20 INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS	.326	.369
item21 INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING	.516	.549
item22 I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION	.397	.444
item23 COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	.662	.791
item24 COMPARED TO OTHER COURSES THIS COURSE WAS	.526	.632

Extraction Method: Principal Axis Factoring.

Rotated Factor Matrix[a] [b]

	Factor[c]		
	1	2	3
item13 INSTRUC WELL PREPARED	.771		
item14 INSTRUC SCHOLARLY GRASP	.726		
item15 INSTRUCTOR CONFIDENCE	.676		
item16 INSTRUCTOR FOCUS LECTURES	.591		
item17 INSTRUCTOR USES CLEAR RELEVANT EXAMPLES	.587	.446	
item18 INSTRUCTOR SENSITIVE TO STUDENTS		.739	
item19 INSTRUCTOR ALLOWS ME TO ASK QUESTIONS		.727	
item20 INSTRUCTOR IS ACCESSIBLE TO STUDENTS OUTSIDE CLASS		.540	
item21 INSTRUCTOR AWARE OF STUDENTS UNDERSTANDING	.402	.533	.321
item22 I AM SATISFIED WITH STUDENT PERFORMANCE EVALUATION		.559	
item23 COMPARED TO OTHER INSTRUCTORS, THIS INSTRUCTOR IS	.449	.377	.668
item24 COMPARED TO OTHER COURSES THIS COURSE WAS	.324	.321	.652

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

B3. An industrial engineer is investigating the effect of four methods of assembly (referred to as M1, M2, M3 and M4) on the assembly time for a color television component. Four operators, Ramlal, Shyamlal, Jadukrishna and Madhu are selected for the study. The engineer knows that each assembly method produces such fatigue that the time required for the last assembly may be greater than the time required for the first, regardless of the method. To account for this source of variability, the engineer uses randomization as shown in the design below.

Order of Assembly	**Operator**			
	Ramlal	Shyamlal	Jadukrishna	Madhu
1^{st}	M3=10	M4=14	M1=7	M2=8
2^{nd}	M2=7	M3=18	M4=11	M1=8
3^{rd}	M1=5	M2=10	M3=11	M4=9
4^{th}	M4=10	M1=10	M2=12	M3=14

Table 2: Experimental Data for Question B3.

(a) What sort of experimental design is being used here?

(b) What is the objective of the study? What is the null hypothesis? (Hint: *treatment*)

(c) What would have been the drawback, if any, of using a two-factor RCD model? Ignoring the order of assembly, what would have been the research conclusion with this experimental data?

(d) Analyze the data from this experiment (α =0.05) and draw appropriate conclusions.

Section C

This is a compulsory question carrying 15 marks. **(1 x 15)**

C1. This uses 1995 economic data where 101 countries of the world were surveyed. This question uses 7 data items from this set tabulated herein. We attempt to create a regression model for per capita GDP of a country (using a logarithmic transformation), using the 6 other variables as predictors.

Sl.No.	Data Name	Type	Value Levels	
1.	DENSITY	metric		population density
2.	URBAN	metric		urban population percentage
3.	LIFEEXPF	metric		female life expectancy
4.	REGION	factor	Pacific / Asia	
			Latn America	
			OECD	*used as base in dummy variable model*
			Middle East	
			East Europe	
			Africa	
5.	CLIMATE	factor	arid / desert	
			arid	
			tropical	
			mediterranean	
			maritime	
			temperate	
			arctic / temp	
			desert	*used as base in dummy variable model*
6.	DEATH_RT	metric		death rate per 1000 population
7.	LOG_GDP	metric		log of per capita Gross Domestic Product **dependent**

Table 3: World 95 data for 101 countries

The correlation of the metric variables is shown here.

	DENSITY	URBAN	LIFEEXPF	LOG_GDP	DEATH_RT
DENSITY	1.00	0.22	0.12	0.16	-0.12
URBAN	0.22	1.00	0.74	0.75	-0.50
LIFEEXPF	0.12	0.74	1.00	0.83	-0.71
LOG_GDP	0.16	0.75	0.83	1.00	-0.41
DEATH_RT	-0.12	-0.50	-0.71	-0.41	1.00

The first 5 rows of the data set are displayed here.

```
      COUNTRY DENSITY URBAN LIFEEXPF       REGION LOG_GDP   CLIMATE DEATH_RT
1 Afghanistan    25.0    18       44 Pacific/Asia   2.312      arid       22
2 Argentina      12.0    86       75 Latn America   3.532 temperate        9
4 Australia       2.3    85       80         OECD   4.227      arid        8
5 Austria        94.0    58       79         OECD   4.265 temperate       11
6 Azerbaijan     86.0    54       75  Middle East   3.477      arid        7
```

The ANOVA table and the Regression table of Coefficients follows.

```
Analysis of Variance Table

Response: LOG_GDP
          Df Sum Sq Mean Sq F value Pr(>F)
DENSITY    1   1.00    1.00   16.93  9e-05 ***
URBAN      1  21.47   21.47  365.16 <2e-16 ***
LIFEEXPF   1   6.39    6.39  108.62 <2e-16 ***
REGION     5   4.65    0.93   15.81  6e-11 ***
CLIMATE    7   0.74    0.11    1.79  0.099 .
DEATH_RT   1   0.38    0.38    6.52  0.012 *
Residuals 84   4.94    0.06
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:
lm(formula = LOG_GDP ~ ., data = v2)

Residuals:
    Min      1Q  Median      3Q     Max
-0.5435 -0.1017 -0.0042  0.1145  0.6412

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          -2.99e-02   7.25e-01   -0.04  0.96726
DENSITY               6.32e-05   4.45e-05    1.42  0.15867
URBAN                 6.69e-03   1.67e-03    4.00  0.00013 ***
LIFEEXPF              4.76e-02   8.18e-03    5.81  1.1e-07 ***
REGIONEast Europe    -3.16e-01   9.25e-02   -3.42  0.00097 ***
REGIONPacific/Asia   -3.44e-01   1.47e-01   -2.34  0.02172 *
REGIONAfrica         -1.74e-01   1.58e-01   -1.10  0.27415
REGIONMiddle East    -2.43e-01   1.47e-01   -1.66  0.10106
REGIONLatn America   -4.74e-01   1.22e-01   -3.88  0.00020 ***
CLIMATEarid / desert -4.68e-01   1.45e-01   -3.22  0.00183 **
CLIMATEarid          -3.40e-01   1.51e-01   -2.25  0.02680 *
CLIMATEtropical      -4.12e-01   1.36e-01   -3.02  0.00333 **
CLIMATEmediterranean -3.86e-01   1.40e-01   -2.76  0.00716 **
CLIMATEmaritime      -3.54e-01   1.92e-01   -1.84  0.06893 .
CLIMATEtemperate     -4.81e-01   1.41e-01   -3.40  0.00102 **
CLIMATEarctic / temp -3.91e-01   1.90e-01   -2.06  0.04267 *
DEATH_RT              3.85e-02   1.51e-02    2.55  0.01245 *
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.242 on 84 degrees of freedom
Multiple R-squared:  0.875,	Adjusted R-squared:  0.851
F-statistic: 36.8 on 16 and 84 DF,  p-value: <2e-16
```

Questions for this case study *(Question C1.)*

(a) Comment on the model fitment. What is the null hypothesis for the overall *regression model* and how do you conclude from the output whether H_0 is rejected or not?

(b) Write down the linear regression model (using dummy variables). How many separate linear equations do we have in the model?

(c) List down the Regions in descending order of the countries' average per capita GDP. Which Region has the highest per capita GDP; which has the lowest?

(d) List down the Climates in descending order of the countries' average per capita GDP.

(e) Which are the independent variables that are significant at $alpha = .01$?

(f) The independent variable URBAN is significant in the ANOVA model but is insignificant as per the Coefficients table (p-value=0.15867). How would the model be affected if this variable were dropped. How would the new model R^2 and the model significance change?

Memory refreshers; hints

-4.68e-01 $= -4.68 \times 10^{-1} =$ -0.468

$\Pr(> |t|)$ = the observed significance (p-value) of the corresponding 2-tailed t statistic

$\Pr(> F)$ = the observed significance (p-value) of the corresponding F statistic

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \implies SS_{Total} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = (n-1)s^2$$

$$\text{Multiple R squared (Regression)} = \frac{SS_{Regression}}{SS_{Total}}$$

$$\text{Multiple R squared (Regression) also} = \text{Correlation}(Y, \widehat{Y})^2$$

$$\text{F value (Regression)} = \frac{SS_{Regression}/d.f._{Regression}}{SS_{Residual}/d.f._{Residual}} = \frac{MS_{Regression}}{MS_{Regression}}$$

$$\text{ANOVA (overall) model significance} = \frac{\sum_{Factors} SS / \sum_{Factors} df}{SS_{Error}/df_{Error}}$$

F Table (Percentage point of F-Distribution (α=0.05)

Numerator Degrees of Freedom (Numerator DF)

Denominator Degrees of Freedom	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	Inf.
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.10	251.14	252.20	253.25	254.25
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.41
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.85
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.82
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.79
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.74
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.68
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.66
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.63
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.26
Inf.	3.84	3.00	2.61	2.37	2.22	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.40	1.32	1.22	1.05

Figure 1: $F_{\alpha=0.05}$ Table