STATISTICS AND PROBABILITY: AN INTRODUCTORY COURSE

OUTLINE

Session	Date	Topic
1	4/5	Introduction to course; concepts of experimental measurement, samples, data representation.
2	4/7	Histograms, curves, distributions, and cumulative plotting.
3	4/12	Means, medians, mode; variance, sums of squares, standard deviation, skewness, kurtosis.
4	4/26	Central Limit Theorem; introduction to probability concepts.
5	4/28	Further notions of probability, random variables, relationship to distributions - PMF, CDF - joint distributions.
6	5/3	Marginal and conditional properties, moments and expectations; means and variances revisited.
7	5/10	Standard shapes of distributions - Bernoulli, binomial, Poisson.
8	5/12	Hypothesis testing - type I and type II errors - choosing the alternative hypothesis.
9	5/17	More standard distributions - exponential, gamma, normal, lognormal, extreme-value.
10	5/19	Parametric statistical tests, distributions; F, t, and chi-square.
11	5/24	Scatter diagrams, covariance, correlation; notions of cause and effect versus coincidence.
12	6/2	Relationships between variables - linear, log, log-linear, various nonlinear shapes; time-series analysis - a special definition of dependence.
13	6/7	Nonparametric statistics, dependence on size, Kolmogorov-Smirnov test, other useful nonparametric statistics.
14	6/14	Review of SPSS capabilities for data analysis; review of project work and overall course.

The course will be illustrated throughout with examples, based primarily on transit passenger counts, passenger surveys, and other bus system data. Other illustrations will use ideas that are in common usage in everyday life. Those taking the course will be expected to complete homework assignments for each class session, and there will be a Quiz Section once a week to review the homework and to deal with questions.

To execute homework, students will require the use of a calculator, pencil, and paper. No homeworks will be done on the computers at RTD.

The approximate schedule will be:

Tuesdays 7:30 a.m. to 9:30 a.m.

Thursdays 7:30 a.m. to 9:30 a.m., Quiz section 9:30 a.m. to 10:30 a.m.

Additionally, there will be a special project assigned for the week of April 18, and the projects will be evaluated and discussed, together with general questions on the course on June 16 and 21.

PARALLEL READINGS

The text for the course is:

> *Applied Managerial Statistics*, by Park J. Ewart, James S. Ford, and Chi-Yuan Lin, Prentice-Hall, Inc., 1982.

Readings for the course are taken from this text.

SESSION	READINGS
1	Chapters 1 and 2
2	–
3	Chapter 3
4	Chapters 4 and 5
5	Chapters 6 and 7
6	Chapter 8
7	Chapter 9
8	Chapters 10 and 11
9	Chapter 12
10	Chapter 16
11	–
12	Chapter 17
13	Chapters 18 and 19
14	–

It is also recommended that as many of the examples in the book be worked in addition to the exercises set during the course. The examples in the book are to be self-marked — answers are to be found at the end of the book.

SESSION 1 EXERCISES

1. In mid-1985, the RTD will no longer be provided with the revenues from Proposition A that make up the deficit of maintaining a 50 cent fare. Suppose that it is now January 1985, and you are responsible for recommending to the General Manager what fare should be charged from mid-June. A Board decision has been scheduled within 60 days on the new fare. However, there is a referendum for a further half-cent sales tax increase to provide a continuing subsidy to bus operators in Los Angeles County, and this referendum will be voted on in 90 days.

Should the referendum pass, the fares will be able to be maintained at their current 50 cents for at least another year, with the consequent maintenance or improvement of patronage throughout the system. If the fare is increased to $1.00, as has been determined to be required with no local sales tax subsidy, several lines will have to be cut back or eliminated, and patronage will decline.

(a) Prepare a table showing the alternative courses of action, possible states of nature, and possible consequences for the RTD.

(b) What is the major source of uncertainty in this situation?

(c) What is the population of concern in this decision situation?

(d) What is the decision parameter of concern in this problem?

(e) How would you obtain additional information to reduce the amount of uncertainty in the decision, and what information would you seek?

2. Define the scales of measurement of the following items:

- Distance to a bus stop in blocks
- The Richter Scale of earthquake intensity
- Income in dollars per year
- Ethnic origin
- Rating of the "readability" of RTD schedules
- Ratings of movies from G to X
- Trip purpose
- Ratings of RTD service between very favorable and very unfavorable

- Traffic Analysis Zone Numbers

- Men's shoe sizes

3. For each of the measurements of question 2, indicate which ones are continuous and which are discrete, and which ones are approximate and which ones are exact.

SESSION 1 EXERCISES

1. In mid-1985, the RTD will no longer be provided with t
Proposition A that make up the deficit of maintaining a 5
that it is now January 1985, and you are responsible for
General Manager what fare should be charged from mid-June
been scheduled within 60 days on the new fare. However,
for a further half-cent sales tax increase to provide a c
bus operators in Los Angeles County, and this referendum
days.

Should the referendum pass, the fares will be able to be
current 50 cents for at least another year, with the cons
improvement of patronage throughout the system. If the i
$1.00, as has been determined to be required with no loca
several lines will have to be cut back or eliminated, and
decline.

(a) Prepare a table showing the alternative courses of ac
of nature, and possible consequences for the RTD.

(b) What is the major source of uncertainty in this situa

(c) What is the population of concern in this decision s

(d) What is the decision parameter of concern in this pro

(e) How would you obtain additional information to reduc
uncertainty in the decision, and what information would

2. Define the scales of measurement of the following ite

- Distance to a bus stop in blocks
- The Richter Scale of earthquake intensity
- Income in dollars per year
- Ethnic origin
- Rating of the "readability" of RTD schedules
- Ratings of movies from G to X
- Trip purpose
- Ratings of RTD service between very favorable

- Traffic Analysis Zone Numbers

- Men's shoe sizes

3. For each of the measurements of question 2, indicate which ones are continuous and which are discrete, and which ones are approximate and which ones are exact.

Session 1 Exercises — Answers

1.(a)

States of Nature	Alternative Courses of Action: Increase Fare to $1.00	Maintain 50¢ F
Referendum passes	Patronage Loss - Bad PR	Status Quo
Referendum does not pass	Appropriate initial action	Revenue Loss

(b) Voting on the referendum

(c) All voters in Los Angeles County

(d) Number of voters who vote for the referendum and number who against it.

(e) From a sample of Los Angeles County residents, determine:

(i) If the vote were today, what would they vote on the refere

(ii) Are they now registered voters, or will they be by the t of the election?

(iii) Did they vote in the last three elections?

(iv) How did they vote on Prop A in 1982?

RTD

Working Notes

Subtask Date Page of

Subject Preparer

2 & 3

a) Distance to bus stop in blocks	RATIO	DISCRETE	EXACT
b) Richter Scale	RATIO	CONTINUOUS	APPROXIMATE
c) Income in $/year	RATIO	CONTINUOUS	APPROXIMATE
d) Ethnic Origin	NOMINAL	DISCRETE	EXACT
e) Readability Ratings	ORDINAL	DISCRETE	EXACT
f) Movie Ratings	ORDINAL	DISCRETE	EXACT
g) TRIP PURPOSE	NOMINAL	DISCRETE	EXACT
h) RTD Service Ratings	ORDINAL	DISCRETE	EXACT
i) TAZ numbers	NOMINAL	DISCRETE	EXACT
j) MEN'S Shoe Sizes	INTERVAL	DISCRETE	EXACT

RTD 44-9
EFF 1/82
Southern California Rapid Transit District, 425 S. Main St., Los Angeles, CA 90013

SESSION 2 EXERCISES

1. The data in Table 37 (attached) are from a recent RTD bus survey. What is the type of scale used for measurement for this question? What form of graphical presentation of the data would be most useful for these data to provide a rapid summary of the information obtained? Develop the presentation(s) that you feel to be the most useful.

2. Table 24 (attached) represents another part of the RTD On-Board data. Plot a cumulative distribution of the overall data and of any five bus lines. What can you learn from these distributions? What method did you use to select the five bus lines, and why? What other presentations of the data would be useful to you and how would you use them?

3. The third set of data attached (Table A) are fares from the ongoing fare counting program of RTD. These data are collected by one-way trip for a sample of bus runs for part of the day. The period covered includes one peak and the midday. Plot histograms for each bus line in the data set and for the entire sample in Table A. Why would you consider a cumulative distribution not to be useful? For one line of your choosing, plot histograms for each trip. What conclusions can you draw from this?

TABLE 37
REASON FOR NOT USING RTD PASS
BY BUS LINE

Bus Line	Don't Ride Enough	Can't Afford Pass	Don't Know Where To Buy Pass	No Convenient Outlet	Might Lose Pass	Other	Total	Number of Respondents
12	54.5%	15.9%	2.3%	4.5%	13.6%	9.1%	99.90%	44
18	30.0	20.0	15.0	5.0	5.0	25.0	100.00	20
29	18.9	56.8	–	–	16.2	8.1	100.00	37
32	33.3	45.2	2.4	9.5	4.8	4.8	100.00	42
44	37.0	31.5	6.5	10.2	6.5	8.3	100.00	108
47	43.2	29.7	2.7	5.4	16.2	2.7	99.90	37
73	58.3	8.3	16.7	8.3	–	8.3	99.90	12
81	63.8	6.9	12.1	6.9	3.4	6.9	100.00	58
86	36.6	19.5	1.2	8.5	11.0	23.2	100.00	82
88	48.8	22.0	14.6	7.3	–	7.3	100.00	41
89	41.4	25.7	4.3	11.4	4.3	12.9	100.00	70
91	48.4	17.2	7.8	3.1	6.3	17.2	100.00	64
96	41.7	33.3	8.3	–	8.3	8.3	99.90	12
114	52.8	18.5	4.6	4.6	6.5	13.0	100.00	108
152	57.1	7.1	5.4	16.1	1.8	12.5	100.00	56
155	71.4	7.1	7.1	–	–	14.3	99.90	14
156	54.7	7.5	24.5	9.4	1.9	1.9	99.90	53
157	50.0	13.2	10.3	11.8	4.4	10.3	100.00	68
160	50.0	6.3	–	25.0	12.5	6.3	100.10	16
164	57.1	8.6	2.9	11.4	5.7	14.3	100.00	35
165	56.7	16.7	16.7	10.0	–	–	100.10	30
166	54.3	8.6	8.6	2.9	11.4	14.3	100.10	35
168	61.5	7.7	7.7	11.5	–	11.5	99.90	26
169	45.0	15.0	12.5	7.5	5.0	15.0	100.00	80
175	64.6	8.3	8.3	10.4	4.2	4.2	100.00	48
210	48.4	18.8	7.8	7.8	10.9	6.3	100.00	64
354	40.0	6.7	–	6.7	20.0	26.7	100.10	15
424	40.0	36.7	3.3	10.0	3.3	6.7	100.00	30
425	56.3	16.9	2.8	8.5	4.2	11.3	100.00	71
431	62.0	18.0	12.0	2.0	–	6.0	100.00	50
435	62.5	17.2	6.3	7.8	3.1	3.1	100.00	64
451	60.0	13.3	6.7	13.3	–	6.7	100.00	30
452	66.7	22.2	–	–	5.6	5.6	100.10	18
453	33.3	50.0	–	–	–	16.7	100.00	6
454	35.0	25.0	25.0	10.0	5.0	–	100.00	20
484	53.8	19.2	11.5	11.5	–	3.8	99.80	26
488	71.4	12.2	8.2	4.1	2.0	2.0	99.90	49
813	73.1	11.5	–	3.8	11.5	–	99.90	26
821	47.4	5.3	21.1	10.5	5.3	10.5	100.10	19
822	45.2	16.1	6.5	19.4	9.7	3.2	100.10	31
826	45.8	22.9	4.2	6.3	10.4	10.4	100.00	48
831	47.8	17.4	8.7	13.0	8.7	4.3	99.90	23
840	53.6	19.6	7.1	7.1	7.1	5.4	99.90	56
844	42.2	9.4	10.9	17.2	4.7	15.6	100.00	64
846	57.1	15.1	3.4	8.4	5.0	10.9	99.90	119
861	67.9	12.5	1.8	8.9	1.8	7.1	100.00	56
867	36.1	16.7	11.1	11.1	5.6	19.4	100.00	36
869	62.9	12.9	4.3	7.1	2.9	10.0	100.10	70
871	57.7	9.3	9.3	10.3	5.2	8.2	100.00	97
872	38.9	16.7	11.1	16.7	11.1	5.6	100.10	18
OVER-ALL	46.3%	22.8%	6.5%	7.1%	7.4%	9.9%	100.00%	2302

Response Rate: 78.7% (of respondents paying cash fares)

TABLE 24
NUMBER OF CARS
BY BUS LINE

Number of Cars Per Household

Bus Line	None	One	Two	Three	Four	Five or More	Total	Mean No. of Cars/HH	Number of Respon-dents
12	23.4%	34.6%	29.9%	7.5%	2.8%	1.9%	100.0%	1.39	107
18	29.9	31.3	16.4	6.0	10.4	6.0	100.0	1.60	67
29	49.5	28.8	7.2	11.7	1.8	.9	99.9	.92	111
32	36.9	23.8	20.2	9.5	6.0	3.6	100.0	1.39	84
44	38.8	31.8	20.7	6.6	.8	1.2	99.9	1.03	242
47	34.2	30.1	16.4	6.8	8.2	4.1	99.8	1.37	73
73	18.3	33.3	28.3	10.0	3.3	6.7	99.9	1.70	60
81	31.4	22.9	24.5	10.1	7.4	3.7	100.0	1.53	188
86	29.0	39.5	21..0	6.8	3.7	*	100.0	1.17	162
88	27.1	32.3	24.0	11.5	3.1	2.0	100.0	1.39	96
89	52.2	29.7	12.1	1.7	2.6	1.7	100.0	.78	232
91	36.2	38.8	17.1	5.3	2.0	.7	100.1	1.01	152
96	*	*	*	*	*	*	*	*	22
114	20.7	33.3	25.7	11.3	4.1	5.0	100.1	1.63	222
152	24.3	29.0	29.0	7.5	7.5	2.8	100.1	1.58	107
155	*	*	*	*	*	*	*	*	38
156	11.3	25.8	37.9	19.4	3.2	2.4	100.0	1.85	124
157	18.7	23.9	35.8	14.9	3.7	2.8	99.8	1.75	134
160	21.3	24.6	27.9	11.5	13.1	1.6	100.0	1.77	61
164	32.3	32.3	19.4	9.7	3.2	3.3	100.2	1.30	93
165	23.1	35.9	24.4	6.4	9.0	1.3	100.1	1.46	78
166	26.0	32.5	19.5	15.6	3.9	2.6	100.1	1.49	77
168	8.5	28.8	35.6	15.3	6.8	5.1	100.1	1.98	59
169	20.9	30.8	29.7	11.0	4.7	2.9	100.0	1.58	172
175	19.0	37.0	25.0	12.0	5.0	2.0	100.0	1.54	100
210	34.0	33.5	23.1	6.1	1.9	1.4	100.0	1.14	212
354	24.5	32.7	26.5	6.1	6.1	4.0	99.9	1.51	49
424	25.6	22.1	31.4	11.6	4.7	4.7	100.1	1.62	86
425	29.3	31.6	23.1	11.1	3.6	1.3	100.0	1.33	225
431	25.7	26.7	29.7	9.9	5.9	2.0	99.9	1.50	101
435	20.4	31.5	25.9	17.9	1.9	2.4	100.0	1.59	162
451	29.0	21.0	27.4	11.3	6.5	4.8	100.0	1.61	62
452	*	*	*	*	*	*	*	*	29
453	*	*	*	*	*	*	*	*	36
454	27.3	28.8	31.8	9.1	*	3.0	100.0	1.36	66
484	33.9	30.4	23.2	8.9	1.8	1.8	100.0	1.21	56
488	18.2	44.1	25.2	9.1	2.8	.7	100.1	1.36	143
813	25.0	33.8	29.4	4.4	4.4	2.9	99.9	1.38	68
821	*	*	*	*	*	*	*	*	39
822	30.2	31.7	22.2	6.3	3.2	6.4	100.0	1.43	63
826	36.4	36.4	16.9	6.5	1.3	2.6	100.1	1.09	77
831	23.3	33.3	16.7	16.7	1.7	8.4	100.1	1.68	60
840	24.8	33.0	27.5	8.3	2.8	3.6	100.3	1.44	109
844	22.6	30.3	29.0	11.6	3.2	3.2	99.9	1.56	155
846	28.0	25.7	26.6	13.1	2.8	3.7	99.9	1.50	214
861	26.7	34.1	21.5	13.3	1.5	2.9	100.0	1.40	135
867	24.1	34.5	28.7	6.9	3.4	2.2	99.8	1.40	87
869	14.8	31.0	29.0	13.5	5.8	5.8	99.9	1.86	155
871	24.0	37.0	24.0	8.9	2.6	3.6	100.1	1.41	192
872	38.6	31.6	21.1	8.8	*	*	100.1	1.00	57
Overall	33.6%	32.3%	20.9%	8.0%	3.2%	2.0%	100.0%	1.22	5500

*Sample size too small to allow valid statistical comparison

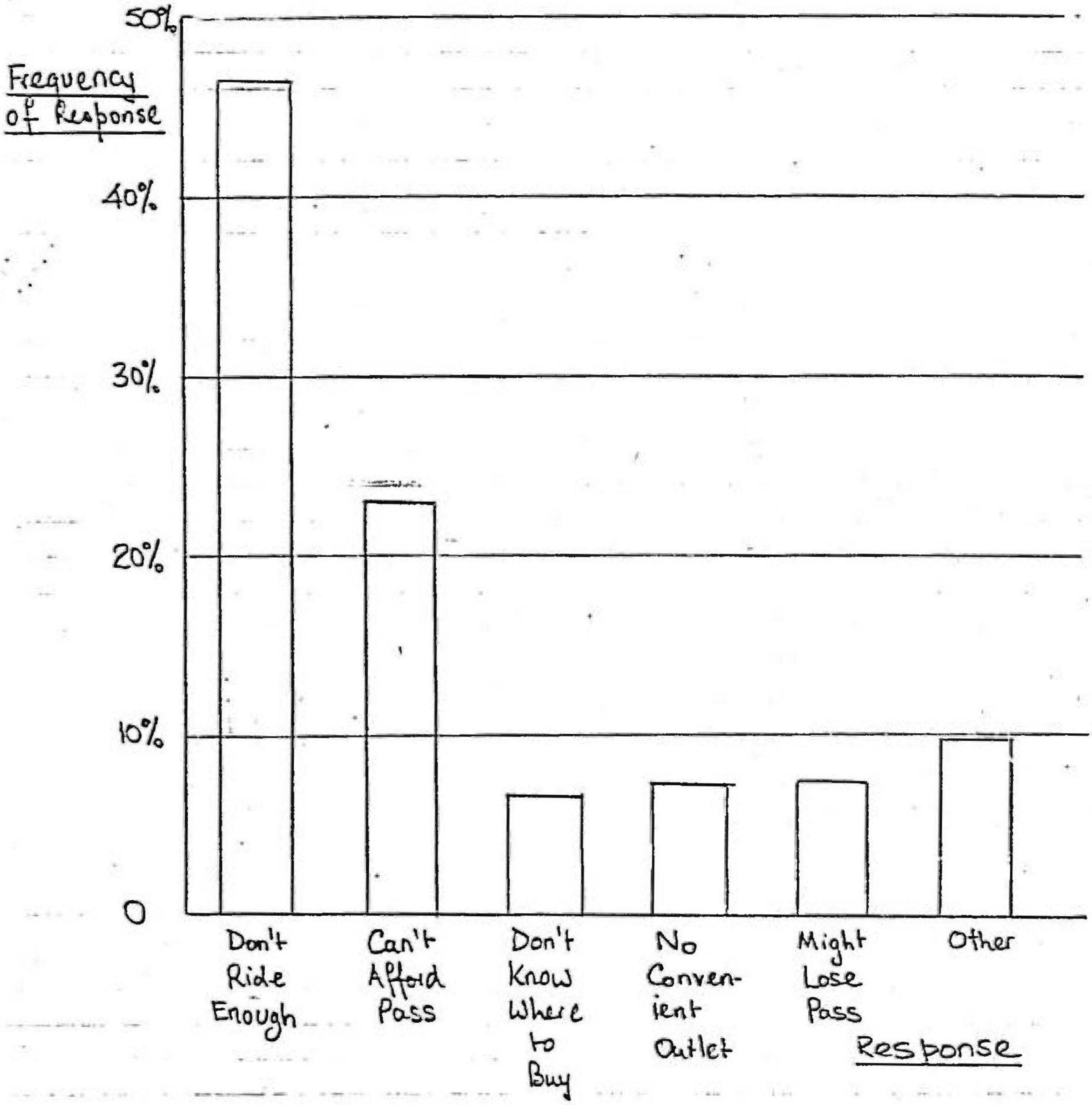
Q2P3

TABLE A

LINE	RUN	TRIP	FARES				PASSES				FREE
			TRANS	$.50 CASH	SENIOR	STUD/ COLL.	REG.	SENIOR	STUD.	COLL.	CHILDREN
30	8	1	26	48	5	0	32	37	16	9	19
30	8	2	42	44	3	0	39	27	17	4	17
30	8	3	11	28	1	0	52	11	5	1	7
30	8	4	2	13	0	0	20	6	2	0	0
30	8	5	23	50	2	0	50	15	14	1	9
30	8	6	13	48	1	0	23	2	6	2	6
5	2	1	22	14	0	0	13	3	1	3	0
5	2	2	1	7	0	0	7	0	1	1	0
5	2	3	4	5	0	0	5	1	0	0	0
5	2	4	0	3	0	0	5	1	0	0	0
5	2	5	1	0	0	0	0	0	0	0	0
5	2	6	16	24	0	1	16	2	0	3	2
5	2	7	3	35	0	0	10	1	1	0	1
207	3	1	49	78	6	1	59	23	38	3	12
207	3	2	43	78	8	2	84	31	12	4	6
207	3	3	23	13	1	0	47	14	21	14	4
207	3	4	40	23	0	0	51	10	31	10	3
207	3	5	22	27	2	0	25	8	7	4	1
207	3	6	14	20	0	0	27	4	5	2	0
207	3	7	20	16	1	0	21	3	1	4	0
207	3	8	12	8	0	0	20	1	7	3	0
12	6	1	36	66	5	0	43	36	10	28	0
12	6	2	36	61	3	0	41	19	25	20	0

12	6	3	46	68	4	0	30	31	19	13	0
12	6	4	26	19	6	0	28	17	27	5	0
12	6	5	14	41	2	0	41	11	6	7	0
12	6	6	12	17	1	2	29	10	8	8	0
5	18	1	19	47	1	2	40	17	18	0	11
5	18	2	22	83	5	1	35	8	15	4	6
5	18	3	37	81	0	0	27	8	36	4	16
5	18	4	18	51	0	0	23	4	6	3	7
44	2	1	3	4	0	0	10	4	0	0	0
44	2	2	8	22	0	0	41	7	2	2	0
44	2	3	11	14	1	0	20	7	8	2	1
44	2	4	10	13	1	0	17	9	4	3	1
44	2	5	21	51	4	0	40	32	7	5	4
44	2	6	37	44	13	0	34	47	16	12	7
44	2	7	21	33	4	0	19	34	19	7	3
44	2	8	20	32	6	0	48	44	22	3	5

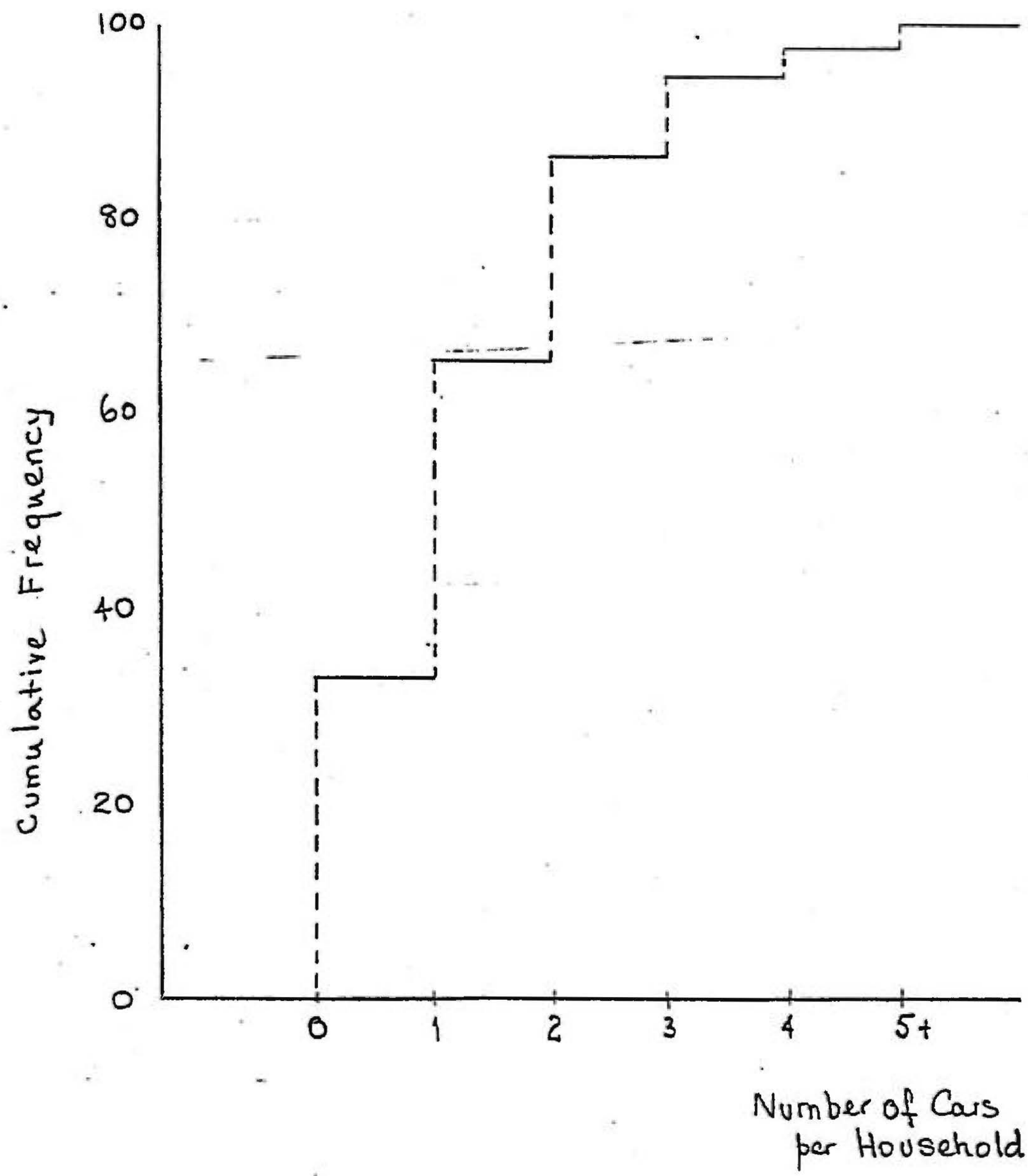
Q.1 a) Scale – Nominal
b) Graphical presentation – Bar Chart of Frequencies – all routes combined



REASON FOR NOT USING RTD PASS

Note: A line-by-line presentation would contain too much information for ready comprehension and there is too much variability in sample size for it to be useful.

Q.2 (a)



(b) Will obviously depend on lines selected. Assume every 10th line is selected, starting at the 4th line (i.e. lines 32, 114, 169, 453, and 844). Because line 453 has no data, we go to the next line with data – line 454 – and count 10 from there for the last line – 846.

Compile the data by obtaining the numbers of responses.

Bus Line	None	One	Two	Three	Four	Five or More	Total
32	31	20	17	8	5	3	84
114	46	74	57	25	9	11	222
169	36	53	51	19	8	5	172
454	18	19	21	6	-	2	66
846	60	55	57	28	6	8	214
Total	191	221	203	86	28	29	758
%	25.2	29.2	26.8	11.3	3.7	3.8	100.0

(Cumulative distribution is on p.4)

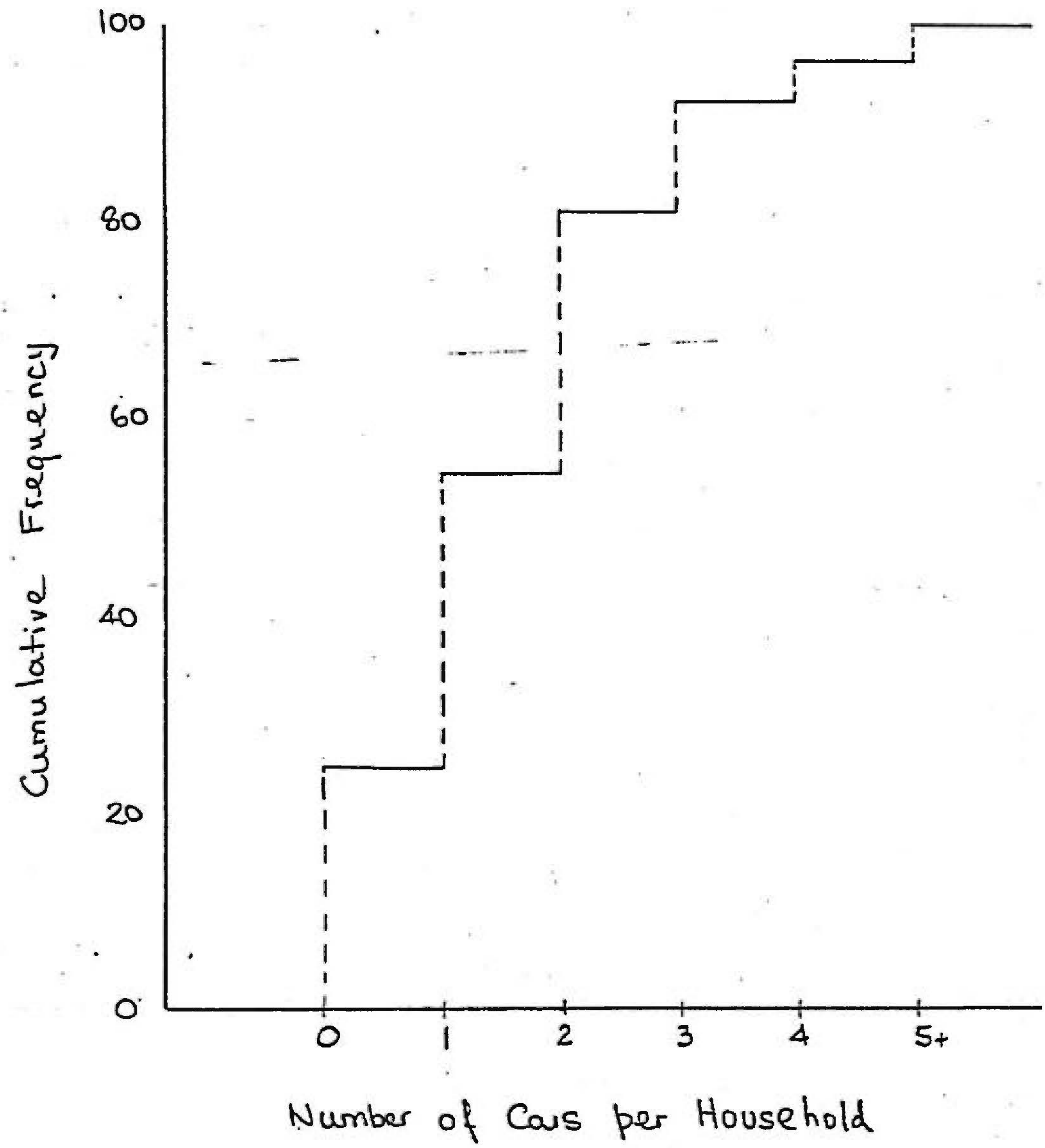
- Selection method is every 10th bus line in the sample, starting from the 4th line, used as a systematic sample.

- Most bus riders own none or one car, and less than 10 percent are from households with three or more cars. The subsample of five lines shows very similar results to the total table.

- A histogram of the frequency distribution would be useful, to show actual auto ownership levels. Also, the distribution of mean cars/household would be useful to show the amount of data variation from line to line.

Subtask Date

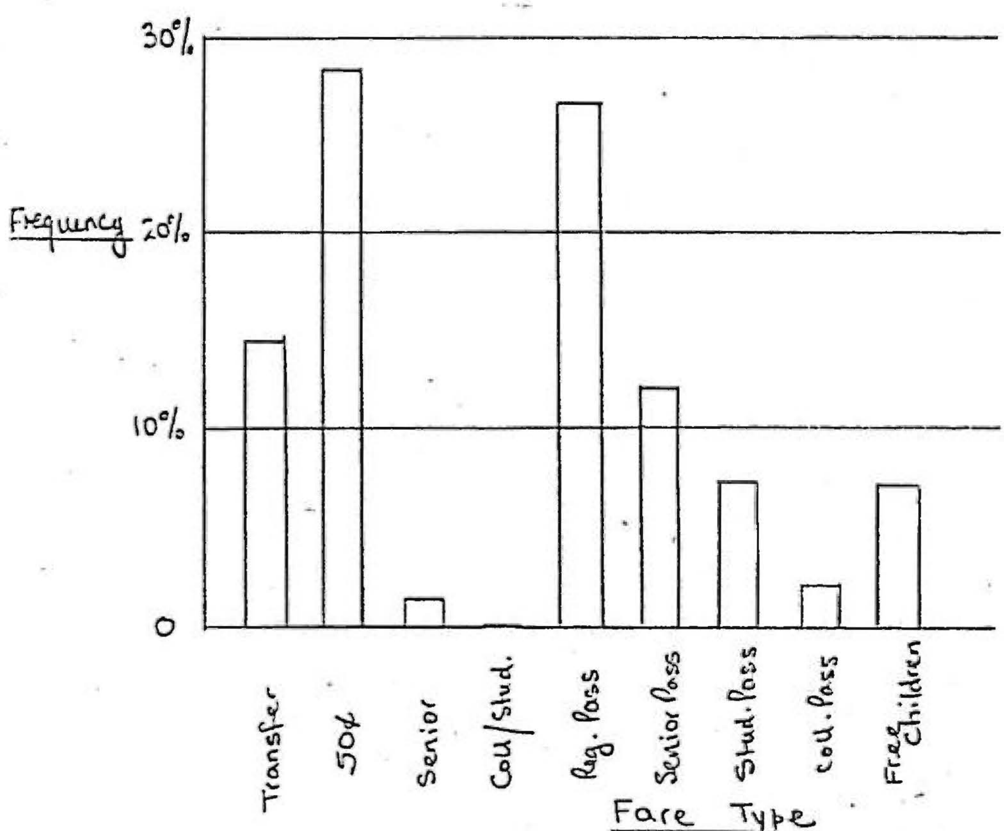
Subject Session 2 Exercises — Answers



Q.3 (a) Line-by-line data are compiled first (n

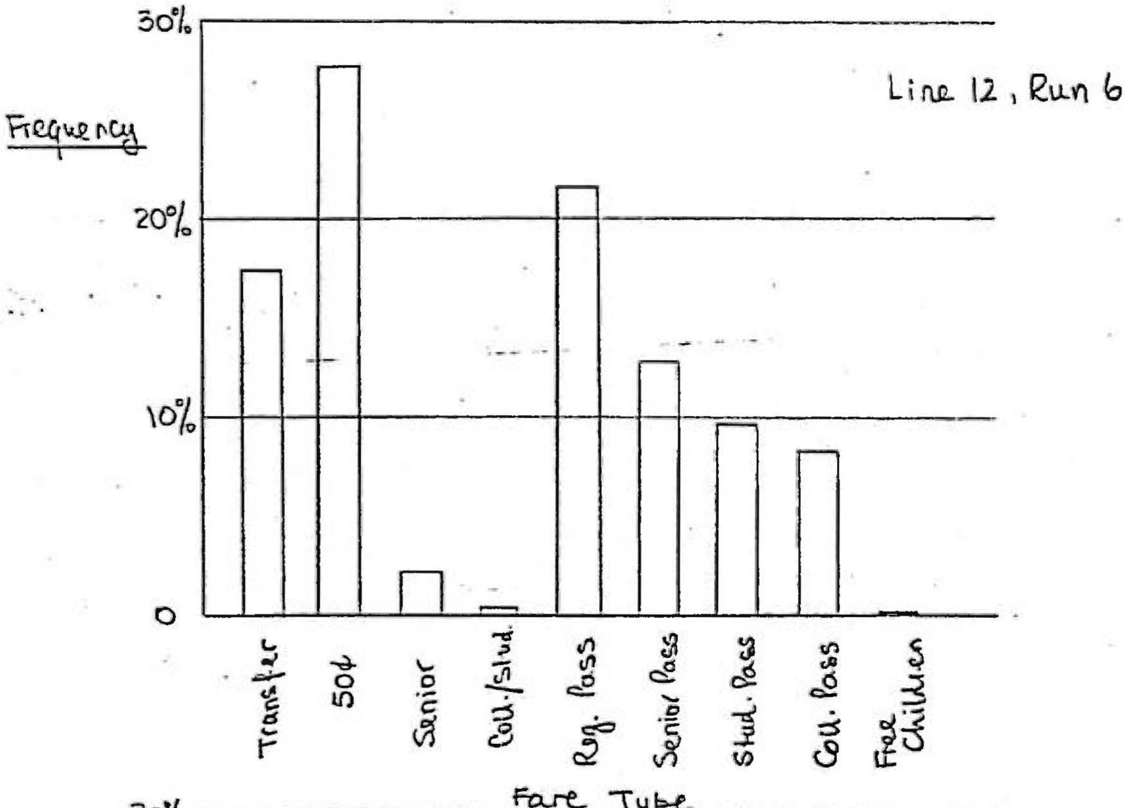
Subtask
Date
Page 5 of 8
Subject: Session 2 Exercises – Answers
Preparer

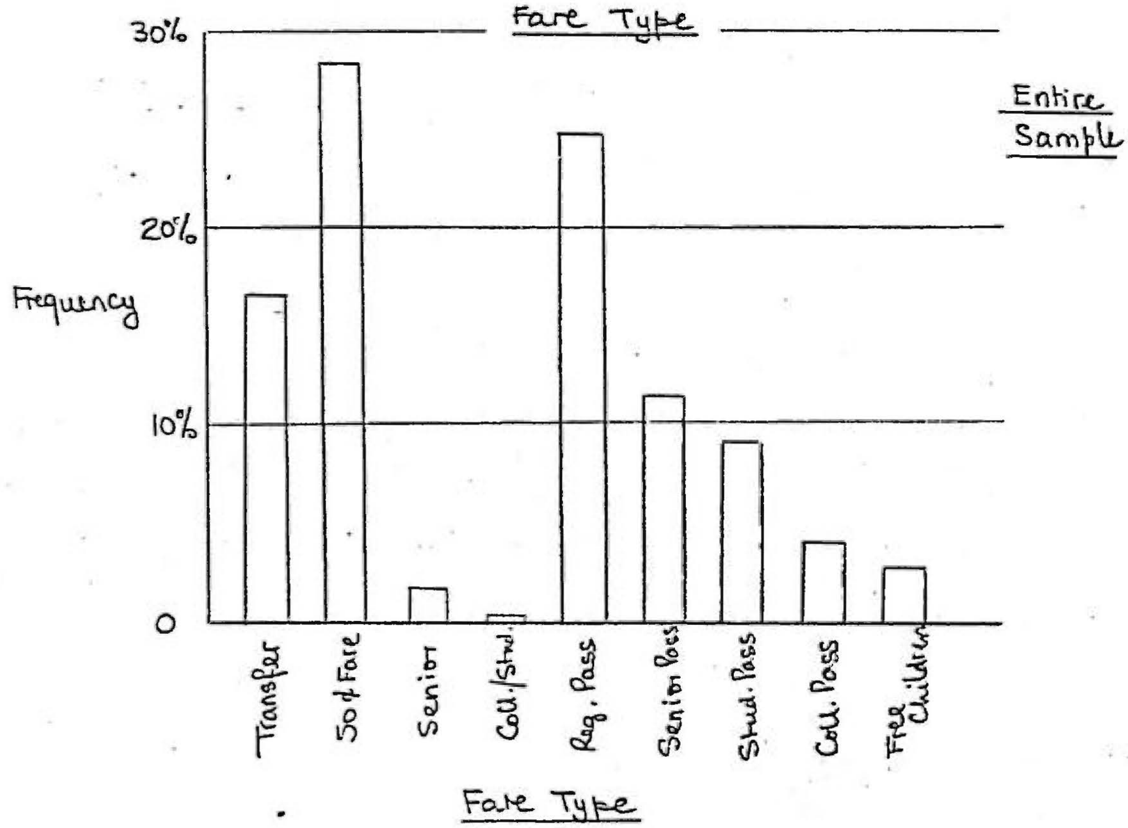
Line	Run	Fares: Trans.	Fares: 50¢	Fares: Senior	Fares: Stud./Coll	Passes: Reg.	Passes: Senior	Passes: Stud.	Passes: Coll.	Free Children	Total
30	8	117	231	12	0	216	98	60	17	58	809
5	2	47	88	0	1	56	8	3	7	3	213
207	3	223	263	18	3	334	94	122	44	26	1127
12	6	170	272	21	2	212	124	95	81	0	977
5	18	96	262	6	3	125	37	75	11	40	655
44	2	131	213	29	0	229	184	78	34	21	919
TOTAL		784	1329	86	9	1172	545	433	194	148	4700



Line 30, Run 8

Histograms should be constructed the same way for the other 5 line/ru entries, e.g., line 12, run 6:

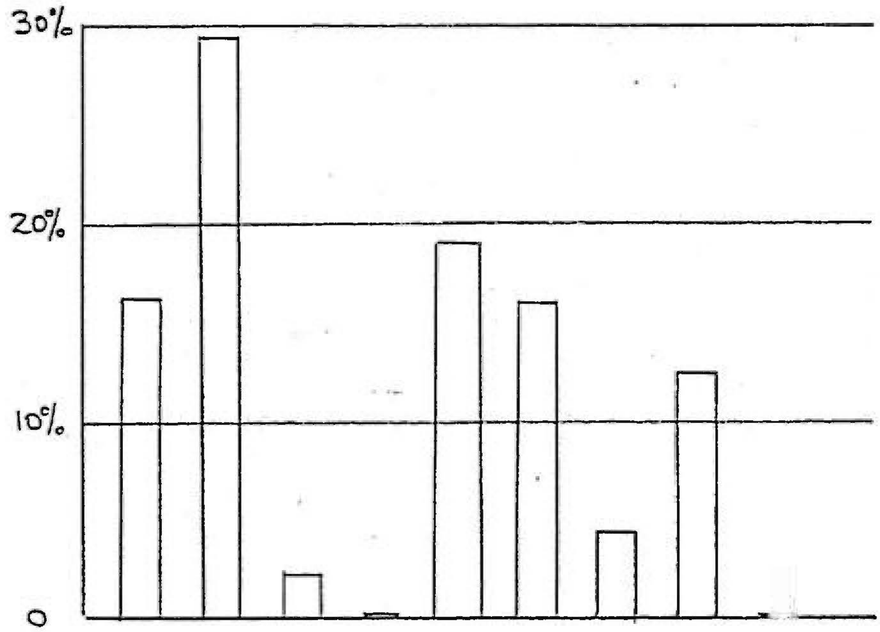




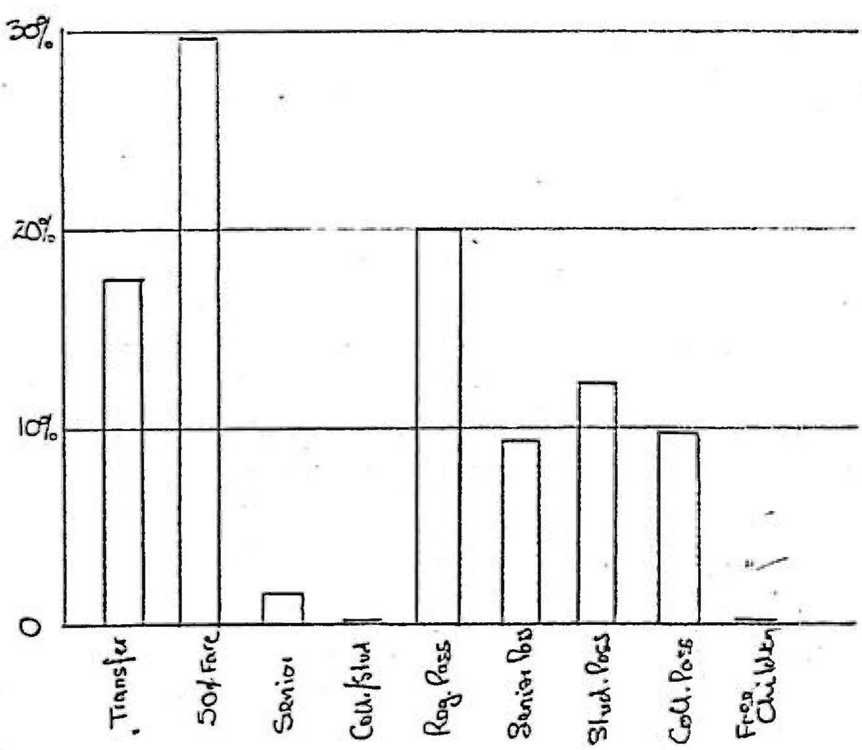
Subject Session 2 Exercises – Answers Preparer

b) Cumulative distribution makes little sense on a nominal scale. There is no specific order of the categories, and any change in order would change the cumulative distribution's shape.

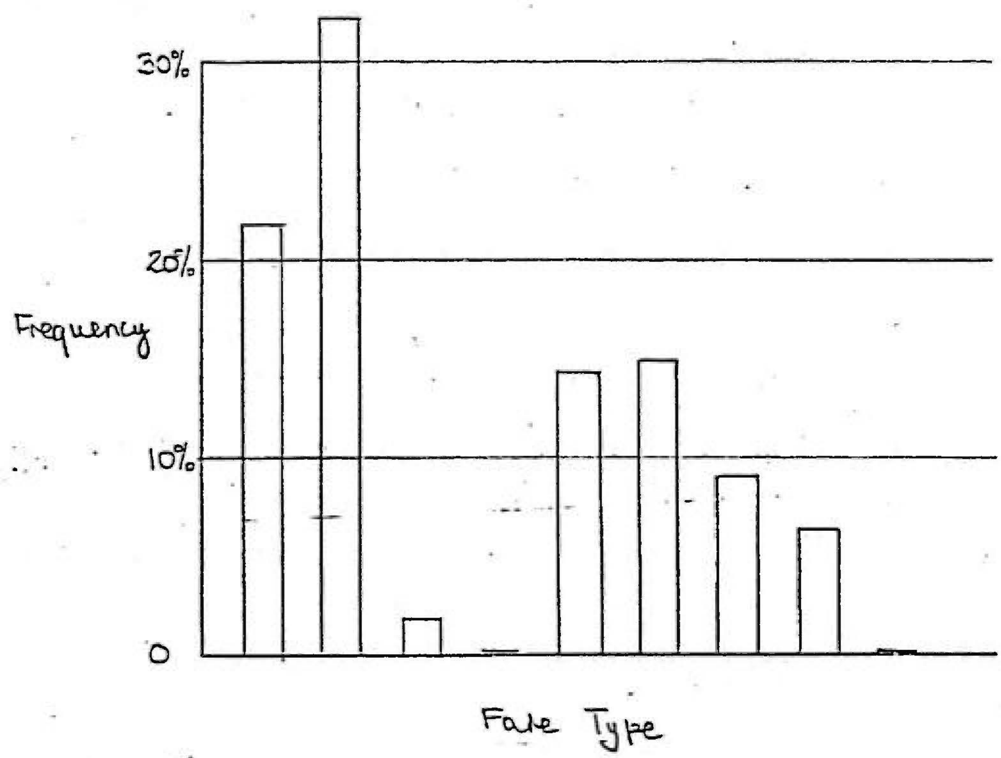
c) Line 12, Run 6 would show something of the following:



Trip 1

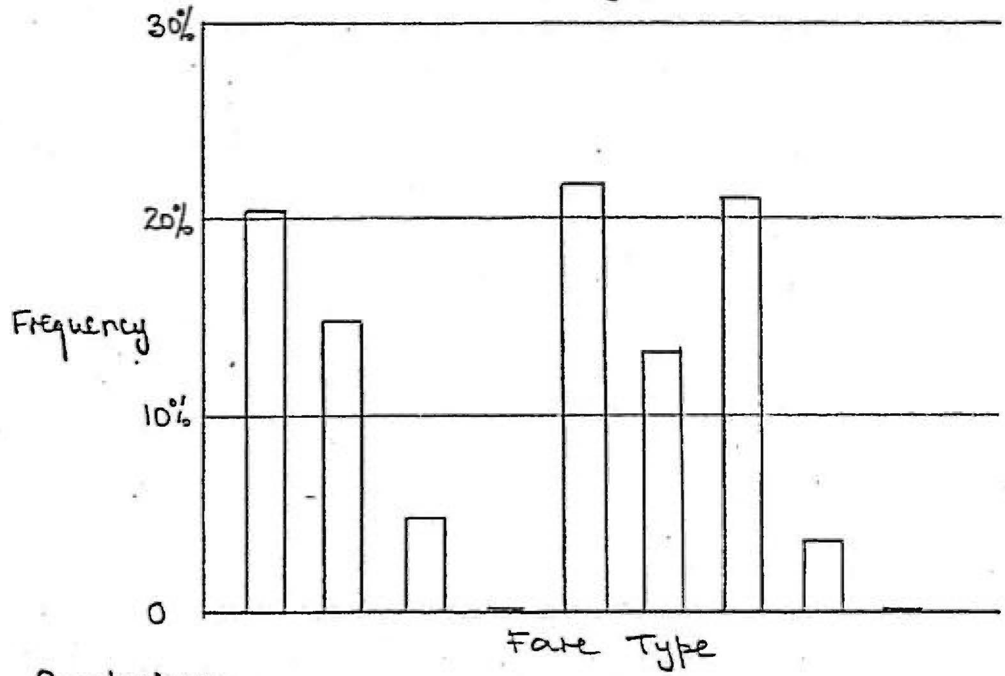


Trip 2



Trip 3

Fare types defined as in previous histograms



Trip 4

Etc.

Conclusions:

a) Average run data hide a lot of variance in fare types

b) For this line, transfers, college/student fares, and free children show little variation from trip to trip; other fare and pass use vary quite markedly

c) Although alternate trips are opposite directions, there is no pattern of directionality in fare use.

SESSION 3 EXERCISES

1. Tables 37 and 24 from Session 2 Exercises are given in perce recompute the tables in terms of absolute numbers of responses f category.

For Table 37, what is the mean number of respondents per bus lin variance and the standard deviation in the numbers of respondent How many respondents do not know where to buy a bus pass? And w average number of respondents per bus line that cannot afford a

2. Calculate the mean, mode, and median car ownership per house survey respondents, from Table 24. Comment on the results of tl calculations. What can you deduce from them? What is the mean respondents to this question per bus line? What is the variance standard deviation, the skewness and the kurtosis? Interpret tl these calculations. Compute the standard deviation of the car c household. Comment on this figure.

3. From Table A in the exercises of Session 2, compute the mear fares and passes for each bus line, and the standard deviation. median for each bus line. Calculate the mean, median, and stan for the entire Table for each fare type. Discuss the results.

RTD **Working Notes**

100 -1 pt wrong

Session 8 Exercises – Answers

Q.1 a) Table 37 becomes:

Bus Line	Don't Ride Enough	Can't Afford Pass	Don't Know Where to Buy Pass	No Convenient Outlet	Might Lose Pass	Other	Total
12	24	7	1	2	6	4	44
18	6	4	3	1	1	5	20
29	7	21	-	-	6	3	37
32	14	19	1	4	2	2	42
44	40	34	7	11	7	9	108
47	16	11	1	2	6	1	37
73	7	1	2	1	-	1	12
81	37	4	7	4	2	4	58
86	30	16	1	7	9	19	82
88	20	9	6	3	-	3	41
89	29	18	3	8	3	9	70
91	31	11	5	2	4	11	64
96	5	4	1	-	1	1	12
114	57	20	5	5	7	14	108
152	32	4	3	9	1	7	56
155	10	1	1	-	-	2	14
156	29	4	13	5	1	1	53
157	34	9	7	8	3	7	68
160	8	1	-	4	2	1	16
164	20	3	1	4	2	5	35
165	17	5	5	3	-	-	30
166	19	3	3	1	4	5	35
168	16	2	2	3	-	3	26
169	36	12	10	6	4	12	80
175	31	4	4	5	2	2	48
210	31	12	5	5	7	4	64
354	6	1	-	1	3	4	15
424	12	11	1	3	1	2	30
425	40	12	2	6	3	8	71
431	31	9	6	1	-	3	50
435	40	11	4	5	2	2	64
451	18	4	2	4	-	2	30
452	12	4	-	-	1	1	18
453	2	3	-	-	-	1	6
454	7	5	5	2	1	-	20
484	14	5	3	3	-	1	26
488	35	6	4	2	1	1	49
813	19	3	-	1	3	-	26
821	9	1	4	2	1	2	19
822	14	5	2	6	3	1	31
826	22	11	2	3	5	5	48
831	11	4	2	3	2	1	23
840	30	11	4	4	4	3	56
844	27	6	7	11	3	10	64
846	68	18	4	10	6	13	119
761	38	7	1	5	1	4	56
867	13	6	4	4	2	7	36
863	44	9	3	5	2	7	70
871	56	9	9	10	5	8	97
872	7	3	2	3	2	1	18
TOTAL	1066	525	150	163	170	228	2302

Southern California Rapid Transit District, 425 S. Main St., Los Angeles, CA 90013

A10P1

RTD

Working Notes

Session 3 Exercises – Answers

Table 24

Bus Line	None	One	Two	Three	Four	Five or More	Total
12	25	37	32	8	3	2	107
18	20	21	11	4	7	4	67
29	55	32	8	13	2	1	111
32	31	20	17	8	5	3	84
44	94	77	50	16	2	3	242
47	25	22	12	5	6	3	73
73	11	20	17	6	2	4	60
81	59	43	46	19	14	7	188
86	47	64	34	11	6	-	162
88	26	31	23	11	3	2	96
83	121	69	28	4	6	4	232
91	55	59	26	8	3	1	152
96	-	-	-	-	-	-	22
114	46	74	57	25	9	11	222
152	26	31	31	8	8	3	107
155	-	-	-	-	-	-	38
156	14	32	47	24	4	3	124
157	25	32	48	20	5	4	134
160	13	15	17	7	8	1	61
164	30	30	18	9	3	3	93
165	18	28	19	5	7	1	78
166	20	25	15	12	3	2	77
168	5	17	21	9	4	3	59
169	36	53	51	19	8	5	172
175	19	37	25	12	5	2	100
210	72	71	49	13	4	3	212
354	12	16	13	3	3	2	49
424	22	19	27	10	4	4	86
425	66	71	52	25	8	3	225
431	26	27	30	10	6	2	101
435	33	51	42	29	3	4	162
451	18	13	17	7	4	3	62
452	-	-	-	-	-	-	29
453	-	-	-	-	-	-	36
454	18	19	21	6	-	2	66
474	19	17	13	5	1	1	56
428	26	63	36	13	4	1	143
813	17	23	20	3	3	2	68
821	-	-	-	-	-	-	39
822	19	20	14	4	2	4	63
826	28	28	13	5	1	2	77
831	14	20	10	10	1	5	60
840	27	36	30	9	3	4	109
844	35	47	45	18	5	5	155
846	60	55	57	28	6	8	214
861	36	46	29	18	2	4	135
862	21	30	25	6	3	2	87
863	25	48	45	21	9	9	155
871	46	71	46	17	5	7	192
872	22	18	12	5	-	-	57
TOTAL	1848	1777	1149	440	176	110	5500

Southern California Rapid Transit District, 425 S. Main St., Los Angeles, CA 90013

A10P2

b) Mean respondents per line = 46 (46.04)

c) Variance $= \frac{1}{n}\sum_i (x_i - \bar{x})^2 = \frac{1}{n}\left(\sum_i x_i^2 - n\bar{x}^2\right)$

$= \frac{1}{50}\left(141854 - 50(46.04)^2\right)$

$= \frac{1}{50} \times 35{,}869.92$

Variance $= \sigma^2 = 717.3984$

Standard Deviation $= \sigma = 26.7843$

d) 150 respondents don't know where to buy a bus pass

e) 10 (11 or 10.5!) respondents on average per bus line cannot affor a pass, out of the average 46 respondents per bus line

Q.2

a) Assuming a car ownership of 5 for the "5 or More" group,

Mean = 1.21 cars / household

Median = 1.0 cars / household

Mode = 0 cars / household

The above values indicate a distribution skewed to the left (because the median and mode are both less than the mean) and that bus riders have generally a lower car ownership than the population in general.

Subject: Session 3 Exercises – Answers

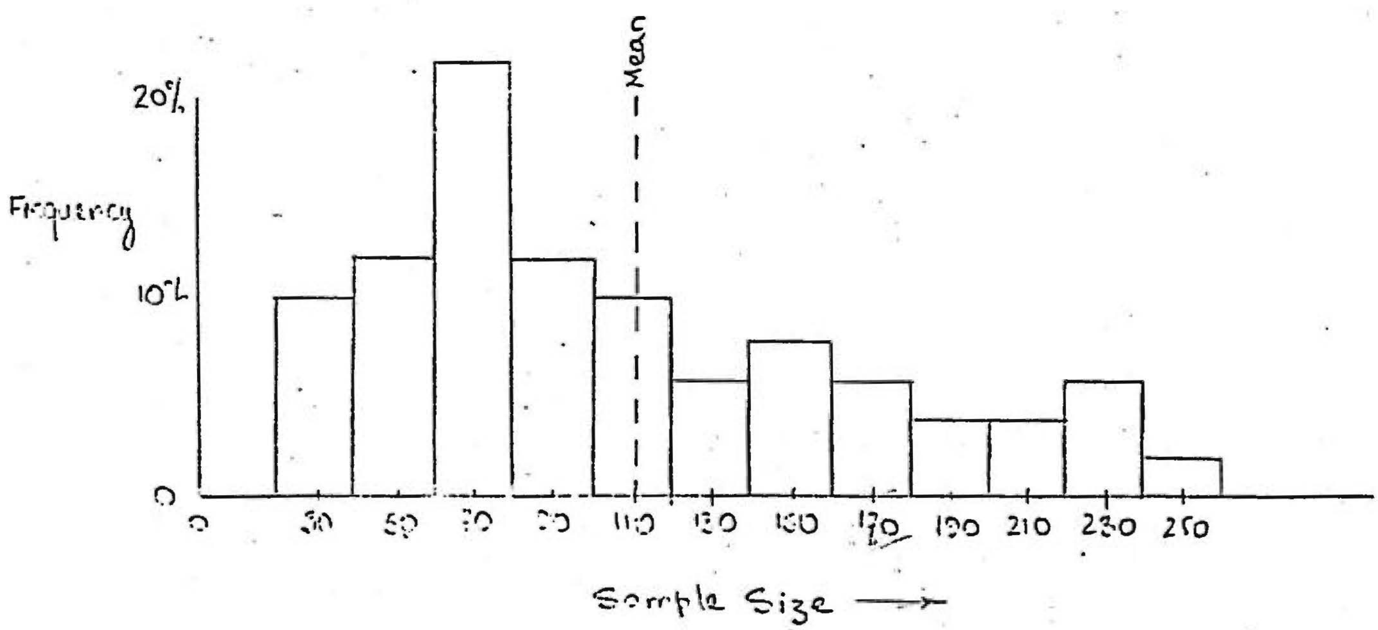
b) Mean respondents per bus line = 110

Variance = 3524.2196

Standard Deviation = 59.365

[Skewness and kurtosis answers to be provided later]

With only variance and standard deviation measured, one can conclude that there is considerable variation in sample size per bus line, because the standard deviation is more than half of the mean. One would also expect, therefore, that few bus lines will have approximately 110 respondents and that the distribution of sample sizes will have a wide range and be fairly flat. To show this, the distribution is provided below, with sample sizes grouped by 20s to show the distribution more clearly.



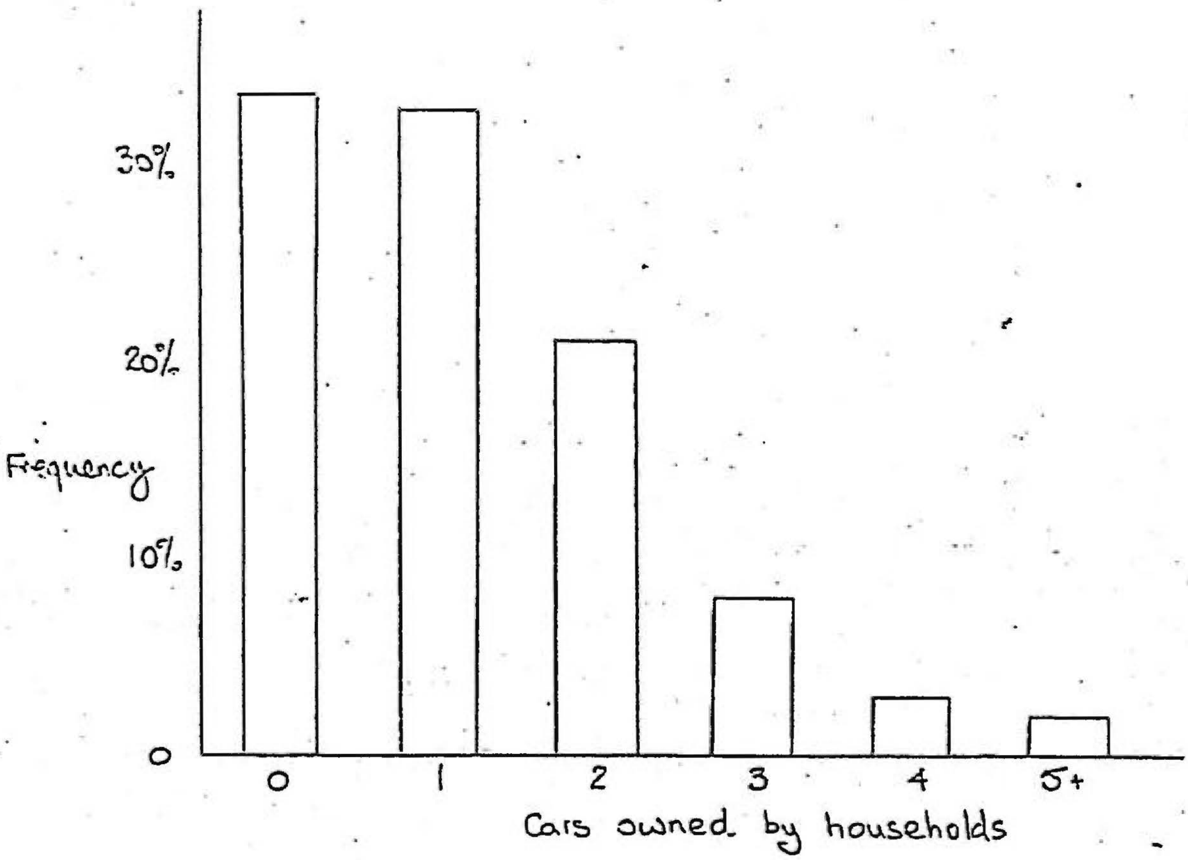
Computation of median (90) and mode (70)
would have indicated the leftward skewing

c) The standard deviation must be computed
not from averages per bus line. The mea
is $\frac{1}{5500}\left[0 \times 1848 + 1 \times 1777 + 2 \times 1149 + \ldots + 5 \times 110\right] = 1.21$

Standard deviation = 1.196

The standard deviation is almost equal t
a high degree of variability in householc
[NOTE: assuming 6 cars as the average for "
increase the mean to 1.23 and the stand
1.265. These small increases do not change
Again, noting the mean, median, and mode
the graphical distribution is shown on the
As further information, plot a histogram of
2.5 (p.24) of the text, and the cumulati
distribution.

Subject: Session 3 Exercises – Answers



Q.3 [NOTE: Answers below do not include transfers. If you included them, your numbers will be different. Make sure your method is right.]

a) Cash fares paid by bus line are:

line 30 231+12 = 243 on 6 trips Mean = 40.5

line 5(2) 88+1 = 89 on 7 trips Mean = 12.7

line 5(18) 262+6+3 = 271 on 4 trips Mean = 67.75

[line 5 (all) 89+271 = 360 on 11 trips Mean = 32.73]

line 207 263+18+3 = 284 on 8 trips Mean = 35.5

line 12 272+21+2 = 295 on 6 trips Mean = 49.17

line 44 213+29 = 242 on 8 trips Mean = 30.25

Passes for each bus line are:

line 30 216+98+60+17 = 391 on 6 trips Mean = 65.17

line 5(2) 56+8+3+7 = 74 on 7 trips Mean = 10.57

line 5(18) 125+37+75+11 = 248 on 4 trips Mean = 62.0

[line 5(all) 74 +248 = 322 on 11 trips Mean = 29.2

line 207 334+94+122+44 = 594 on 8 trips Mean = 74

line 12 212+124+95+81 = 512 on 6 trips Mean = 85

line 44 229+184+78+34 = 525 on 8 trips Mean = 65

Standard Deviations:

line	Fares	Passes
30	14.67	25.69
5(2)	11.94	7.13
5(18)	17.48	15.92
5(all)	30.05	27.14 ✓
207	30.01	40.29 ✓
12	21.18	21.80
44	18.33	34.95

b) Median numbers of fares and passes for each bus line:

Line	Fares	Passes
30	48 (47/49)	74.5 (69/80)
5(2)	7	9
5(18)	66 (51/81)	68.5 (62/75)
5 (all)	25	20
207	21.5 (20/23)	41 (38/44)
12	53.5 (43/64)	85 (77/93)
44	29.5 (22/37)	65.5 (52/79)

c) Mean, Median, Standard Deviation for entire Table A by fare type

	Transfer	50¢	Senior Fare	Stud/Coll.	Reg. Pass	Senior Pass	Stud Pass	Coll. Pass	Chil
Mean	20.1	34.1	2.2	0.23	30.1	13.97	11.1	4.97	3.8
Median	20	28	1	0	28	9	7	3	1
S.D.	13.47	23.70	2.83	0.57	17.05	13.24	10.31	5.77	5.13

d) Discussion

(i) Line-by-line Data

- The number of fare paying passengers varies less than the number of pass-using passengers.
- Although the means of the two line 5 runs differ by 55, their standard deviations are only different by about 5.5.
- There is no consistency in skewness of the fare-paying distributions: lines 30 and 12 are skewed to the right; 5(2) and 207 are skewed right, as is 5(all); and 5(18) and 12 are close to unskewed

A10P8

- Similarly, pass use shows no consistent skewing:

  lines 30 & 5(18) are skewed to the right

  lines 5(all) & 207 are skewed to the left

  lines 5(2), 12, and 44 are not skewed.

- There are about three groups of lines in terms of fare use:

  low – 5(2) medium – 44, 207, 5(all) high – 12, 30, 5(18)

- There are also two groups of lines in terms of pass use:

  low – 5(2), 5(all) high – 30, 44, 5(18), 207, 12

- Standard deviations do not group in the same way as means. Therefore, grouping lines may not reduce standard deviations.

(ii) All Data

- All fare types and pass types are skewed to the left.
- Standard deviations are generally large, being of the same order of magnitude as the mean. In relation to the mean, regular pass use has the lowest variability, student/college fares the highest.
- Based on the means, an average bus trip from the sampled lines would carry approximately 120 passengers, 36 of whom would pay cash fares, yielding farebox revenue of $18.40, 60 would use a pass, 4 would be children travelling free, and 20 will use a transfer.

Subject: Session 3 Exercises — Answers

- Because of the left skew in the distributions, there will be more bus runs that will have less than the mean numbers and fewer that will have more. Using the medians, about half of all bus trips will have up to 97 passengers, farebox revenue of $14.20, 47 pass users, 1 free child, and 20 transfers.

SESSION 4 EXERCISES

1. The RTD bus fleet consists of (say) 327 Grumman buses, 256 new GM buses, 763 older GMC buses, 317 AM Generals, and 449 assorted other buses. Assume that the allocation of buses by make to each run of RTD bus routes is a random occurrence, and that your arrival at a particulat bus stop in downtown Los Angeles is also a random event. What are the probabilities of the following events:

a. The first bus to arrive at your stop is a Grumman? .

b. The first bus to arrive is a new GM, and the second is an AM General?

c. The third bus to arrive is another new GM?

d. The fourth bus to arrive is neither a Grumman, nor a new GM?

2. For the RTD bus fleet of Question 1, suppose that 100 buses were selected randomly. How many of the buses would be expected to be of each type in such a random selection?

Using these numbers, what is the number of permutations of buses by bus type that could be sent out from the garage? And how many combinations are there? What are the odds that the first bus that is sent out from the garage is a new GM bus?

3. Use set theory to define the possible outcomes of tossing a coin three times; and of rolling a pair of dice. Is the set {11,12,13} a proper subset of the second set? Why?

Suppose you are playing Blackjack with a standard pack of 52 cards. You have been dealt a nine and a seven and the dealer has dealt two cards facedown to himself or herself. If you were to ask for one more card, what is the set of possible outcomes that can occur? If the dealer has a King and a 5, what is the universal set of outcomes? What is the probability that your third card will leave you with a score of 21 or less? (Assume that an Ace scores 1 and that picture cards are worth 11.) Suppose that your third card is a 4. What are the odds that the dealer will beat you, if he or she draws one more card? Calculate the odds by using sets.

Subject: Session 4 Exercises - Answers

Q.1

a) Total fleet given is 2112 buses

Probability of first bus being a Grumman is 327 ÷ 2112

= 0.155

b) Probability of a new GM bus arriving is 256 ÷ 2112 = 0.121

Probability of an AM General bus arriving is 317 ÷ 2112 = 0.150

These are two independent events, therefore, probability of bus #1 being a new GM and #2 an AM General

= 0.121 x 0.150 = 0.018

c) Probability of another new GM bus arriving is 255 ÷ 2111 = 0.12

(because 1 new GM bus just went by and is not able to come by again).

The probability of bus #1 being a new GM, #2 an AM General, and #3 a new GM = 0.121 x 0.150 x 0.121

= 0.0022

d) The probability that a bus is neither a Grumman nor a new GM is (2112 - 327 - 256) ÷ 2112 = 0.724

But, given that three buses have already passed, we should calculate this probability as

(2109 - 326 - 254) ÷ 2109 = 0.725

Then, the probability of observing the sequence of: new GM - Grumman - new GM - [not Grumman, not new GM] is $0.0022 \times 0.725 = 0.001595$

Q.2

a) Random selection of 100 buses assumes the same proportion by bus type as in the total fleet. Therefore, composition of the sample of 100 buses should be:

Grumman	– 16
new GM	– 12
GMC	– 36
AM Gen.	– 15
Other	– 21
	100

b) The number of permutations of the 100 buses is

$$_{100}P_{100}(16,12,36,15,21) = \frac{100!}{16!\,12!\,36!\,15!\,21!} = 3.7468682 \times 10^{62}$$

The number of combinations of 100 buses taken together

is $$_{100}C_{100} = \frac{100!}{100!\,1!} = 1$$

c) The odds that the first bus is a new GM is 12/100 to 88/100 or 3:22 or 1 to $7\frac{1}{3}$.

Q.3

a) Set of outcomes of coin tossing = {(H,H,H), (H,H,T), (H,T,H), (T,H,H), (T,T

(T,H,T), (H,T,T), (T,T,T)} (9 outco

Set of outcomes of rolling a pair of dice = {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1

(2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3

(3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5),

(5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6

(6,3), (6,4), (6,5), (6,6)} (36 outcom

b) {11, 12, 13} is not a proper subset, because a value of cannot be obtained from any outcome.

c) (i) Because you do not know the cards that the dealer has, there are 50 possible outcomes, as follows:

Set for 1 more card = {(4×Ace), (4×2), (4×3), (4×4), (4×5), (4

(3×7), (4×8), (3×9), (4×10), (4×J), (4×Q),

(4×K)} for 50 outcomes.

(ii) The universal set of outcomes is defined in terms of the possible combinations of your 9 and 7 with a third card, together with the dealer's King and 5 with or without a third card.

The set of outcomes for you is:

$$P = \{(4 \times Ace), (4\times 2), (4\times 3), (4\times 4), (3\times 5), (4\times 6), (3\times 7), (4\times 8), (3\times 9), (4\times 10), (4\times J), (4\times Q), (3\times K)\} \quad \text{(48 outcomes)}$$

The set of outcomes for the dealer is:

$$D = \{(\text{no card}), (4\times Ace), (4\times 2), (4\times 3), (4\times 4), (3\times 5), (4\times 6), (3\times 7), (4\times 8), (3\times 9), (4\times 10), (4\times J), (4\times Q), (3\times K)\}$$ less the third card you drew (48 outcomes)

$$\mathcal{U} = P \cup D$$

(iii) To score 21 or less requires that your third card is a 5 or less. Given that the dealer has a King and a 5, the probability of scoring 21 or under is $19/48 = 0.396$

(iv) You now have a score of 20. The dealer can beat you only by scoring exactly 21. Given the King and 5, the dealer has a score of 16 and therefore needs a another 5. The set of possible outcomes is:

$$D_1 = \{(4\times Ace), (4\times 2), (4\times 3), (3\times 4), (3\times 5), (4\times 6), (3\times 7), (4\times 8), (3\times 9), (4\times 10), (4\times J), (4\times Q), (3\times K$$

This defines 47 outcomes, only 6 of which allow the dealer to win. (He or she wins by equalling or beating your score) Therefore, the odds are 6 to 47 that the dealer wins, or slightly better than 1 in 8.

SESSION 5 EXERCISES

1. In planning for the future of SCRTD, two events play a major role defining the directions for development of the bus routing and schedul whether or not additional funds will become available in 1985 to repla subsidy currently provided from Prop A, and whether or not UMTA will p requested funding to build Metro- Rail. The possibility of a referenc for added taxes in 1985 for subsidy purposes is considered to be enhan UMTA has already committed to funding Metro-Rail.

Based on a Consultant's study, it is determined that the probability c additional taxes being passed for 1985 is

$$P(A) = 0.45$$

and that the probability that UMTA will make a commitment to fund Metı

$$P(M) = 0.60$$

It is also estimated that the probability that the referendum will pas increased by 50% if UMTA commits funding to Metro-Rail.

(a) Which of the following is a correct statement? Events A and M are

(i) Independent and mutually exclusive

(ii) Independent but not mutually exclusive

(iii) Mutually exclusive but not independent

(iv) Neither mutually exclusive nor independent.

(b) What is the probability that a referendum will pass and that Met will be funded?

(c) What is the probability that the referendum will pass if UMTA fu Metro-Rail?

(d) What is the probability that neither event takes place?

2. The cost of an on-board survey, using surveyors who ride the buse composed of several elements. The costs are \$15 per hour for the tim surveyors, 20 cents per survey handed out, an overhead cost of \$5 per surveyor for supervision, and a fixed cost of \$25 per survey day for administrative and related expenses.

(a) Write down a function that expresses the daily cost of such a sur define the variables you use.

(b) If a particular survey is to involve 40 hours of surveyors per da will hand out an average of 50 forms an hour, and each surveyor will work an average of 6 hours per day, what is the daily cost of the survey?

(c) A second survey firm pays their surveyors on a piece-rate instead of hourly, with a rate of 50 cents per completed and returned survey. This firm's printing costs for the forms is 18 cents per survey, their supervision costs the same, and the fixed administrative charge is $30 per day. Write out the cost function of this firm.

(d) If the first firm achieves a response rate of 55 percent, and the second firm achieves a response rate of 65 percent, which firm will be least expensive under the conditions of (b)?

3. Consider question 6.6 (page 115) in the text. After answering questions (a) and (b), answer (c):

(c) Suppose that the coin is to be tossed five times. Specify the sample space for this experiment. If X denotes the number of heads resulting from the experiment, is X a discrete or a continuous random variable? Define the probability function for this experiment, and specify if it is a probability MASS or a probability DENSITY function.

Q.1

(a) (iii) is the correct statement – they are mutually exclusive but they are not independent.

(b) $P(A) = 0.45$ but increases to 0.675 if UMTA funds Metro-Rai

$$P(A \cup M) = 0.675 \times 0.60 = 0.405$$

(c) The probability that the referendum will pass if UMTA funds Metro-Rail (see (b) above) is $1.5 \times 0.45 = 0.675$

(d) $P(A' \cup M') = 0.55 \times 0.40 = 0.22$

Because $P(M')$ = UMTA does not fund Metro-Rail

$= 1 - 0.60 = 0.40$

$P(A')$ = Referendum does not pass

$= 1 - 0.45 = 0.55$

Q.2

(a) Let x = number of surveyor hours per day, w = number of survey

y = number of surveys handed out per surveyor hour

$f(c) = 15x + 0.20yx + 5x + 25$ $\qquad f(c)$ = daily cost in dollar

$= 20x + 0.20xy + 25$

(b) $f(c) = 20 \times 40 + 0.20 \times 50 \times 40 + 25$

$= \$1225$

A5P1

Subject Session 5 Exercises — Answers

Preparer

(c) $f(C') = 0.18xy + 0.5rxy + 5x + 30$

where x, y are as defined for (a).

r = response rate of bus riders

(d) Under condition (b), the first firm will distribute 2000 surveys per day. With a 55 percent response rate, this means that 1100 surveys will be returned. The cost per returned survey is $1225/1100 = $1.136

Under the same conditions, the second firm's cost is:

$f(C') = 0.18 \times 40 \times 50 + 0.5 \times 0.65 \times 40 \times 50 + 5 \times 40 + 30$

$= \$1240$

This firm will obtain $0.65 \times 2000 = 1300$ returned surveys at an average cost of $1240/1300 = $0.9538

Therefore, the second firm is least expensive.

Alternatively, to hand out sufficient surveys to get 1100 returns (= 1692 surveys — 1100/0.65), at 50 surveys per hour, the second firm will have to work for 34 hours (1692/50 = 33.84 ≈ 34).

$f(C')_{1100} = 0.18 \times 34 \times 50 + 0.5 \times 1100 + 5 \times 34 + 30$

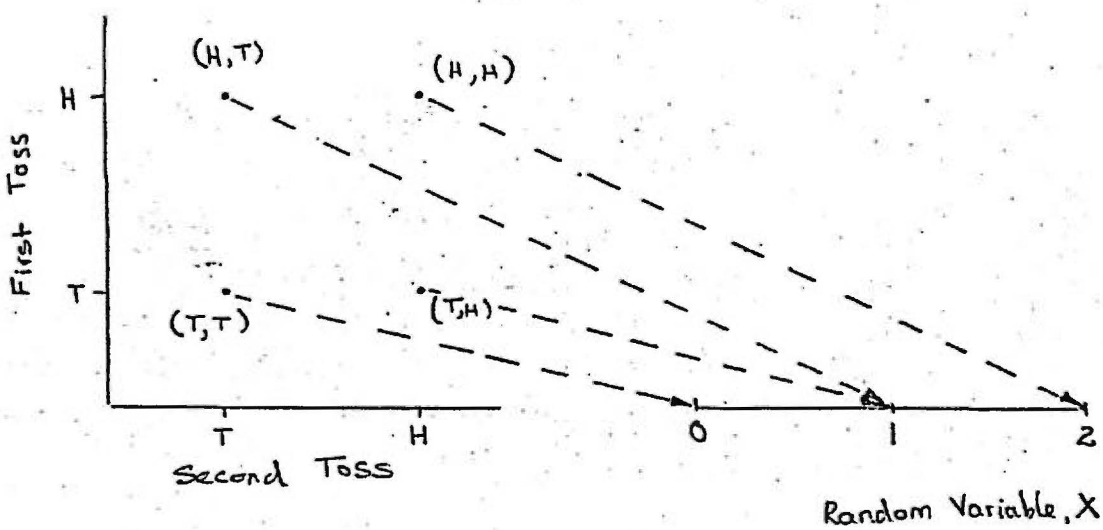
$= \$1056$

Compared to the first firm's $1225, firm 2 is least expensive

Q.3 (6.6)

(a)

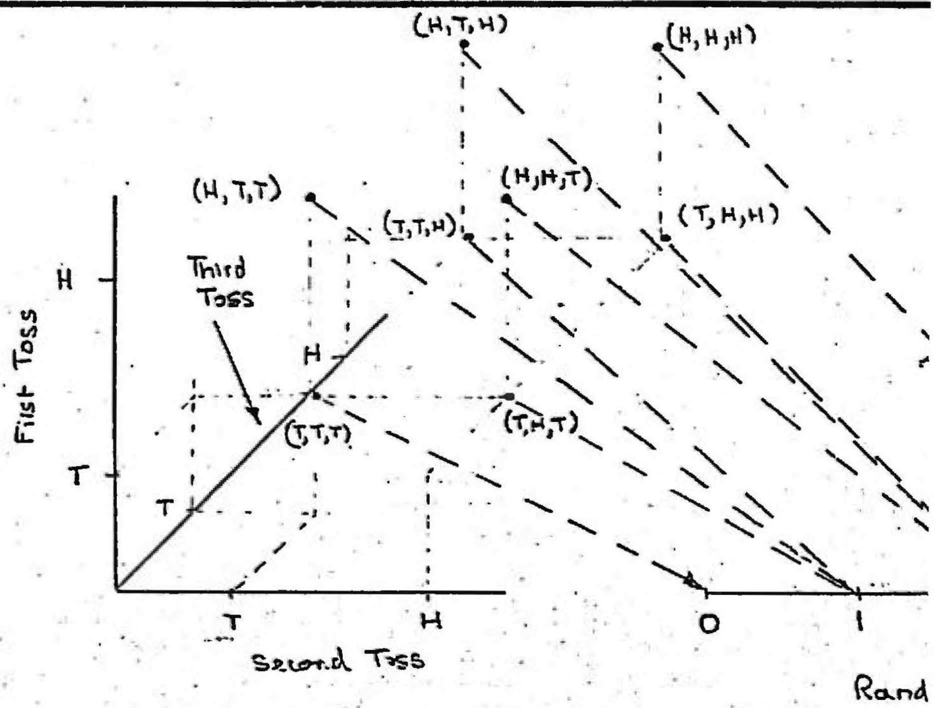
Sample space is {(H,H), (H,T), (T,H), (T,T)}



(b) For tossing the coin three times, the sample space is:

{(H,H,H), (H,H,T), (H,T,H), (T,H,H), (H,T,T), (T,H,T), (T,T,H), (T,T,T)}

The diagramming problem clearly is that a three-dimensional drawing is now required for the sample space (as shown on page 4). This can be resolved, instead, into the second diagram of page 4.



Random Variable Y	Sample Space Outcome
0	(T, T, T)
1	(T,T,H), (T,H,T), (H,T,T)
2	(T,H,H), (H,T,H), (H,H,T)
3	(H,H,H)

(c) For five tosses of the coin, the sample space be

[(H,H,H,H,H), (H,H,H,H,T), (H,H,H,T,H), (H,H,T,H,H),

(H,H,H,T,T), (H,H,T,H,T), (H,T,H,H,T), (T,H,H,H,

(H,T,H,T,H), (H,T,T,H,H), (T,H,H,T,H), (T,H

(T,T,T,H,H), (T,T,H,T,H), (T,H,T,T,H), (H,T,T,

(T,H,T,H,T), (T,H,H,T,T), (H,T,T,H,T), (H,T,

(H,T,T,T,T), (T,H,T,T,T), (T,T,H,T,T), (T,,T,T

(T,T,T,T,T)}

X is a discrete random variable

The PROBABILITY MASS FUNCTION is determined as follows:
The sample space defines 32 outcomes. Each outcome has a probability of 1/32 of occurring. Defining the relationship gives:

Random Variable X	Number of Points	f(x)
0	1	.03125
1	5	.15625
2	10	.3125
3	10	.3125
4	5	.15625
5	1	.03125

(These can be counted from the sample space, or a diagram can be constructed.) Note that the number of points for each value of X is the number of permutations of 5 objects of two types where X are of type 1 (Heads) and Y are of type 2 (Tails), and Y = 5-X. Therefore, we can define

$$f(x) = \begin{cases} \dfrac{5!}{x!\,(5-x)!} \cdot \dfrac{1}{32} & x = 0,1,2,3,4,5 \\ 0 & \text{otherwise} \end{cases}$$

SESSION 6 EXERCISES

1. Let X be a random variable denoting the number of RTD bus breakdowns in one division in a 24-hour period. The probability mass function of X is given by:

$$f(x) = \begin{cases} .04 & \text{if } X = 0 \\ .005(12-X) & \text{if } X = 1,2,3,4,5 \\ .005(X+2) & \text{if } X = 6,7,8,9,10,11,12 \\ .01(21-X) & \text{if } X = 13,14,15,16,17,18,19 \\ 0 \text{ otherwise} & \end{cases}$$

(a) What is the probability that there will be 7 breakdowns in a 24-hour period from one division?

(b) Find the probability that there will be more than 2 breakdowns, but less than 10.

(c) Draw the distribution. Calculate the expected value of X, the mode(s), and the variance of X. Show the position of the mean and the mode(s).

2. Let X be a random variable denoting the length of time (in months) before a RTD employee loses his or her ID badge/bus pass. The probability density function of X is given by:

$$f(x) = \begin{cases} 0.1 & \text{if } 0 \le x < 3 \\ a + bx & \text{if } 3 \le x < 6 \\ c - dx & \text{if } 6 \le x < 9 \\ 0.01 & \text{if } 9 \le x < 12 \\ 0 \text{ otherwise} & \end{cases}$$

(a) Assume that a = 0.05, b = 0.01, c = 0.1358, and d = 0.001. Determine whether these values are consistent by defining the cumulative density function and determining its maximum value.

(b) What is the probability that an employee will lose his or her badge in the 5th month?

(c) What is the expected length of time that an employee will keep
his or her badge?

(d) What is the standard deviation of the time that an employee will keep his or her badge?

(e) If the cost to RTD of replacing the badge is given by

$$C = g(X) = \$.50 + .1(8 + 5x)$$

What is the expected cost of replacing an employee badge? If there are 8,500 employees with badges, what is the expected annual cost of replacing badges?

(f) How many employees (if any) will not lose their badges within 11.5 months? If an employee who loses his or her badge more than two weeks before it expires is charged a replacement fee of $5, irrespective of when in the year he or she loses it, how much profit (loss) will RTD make on replacement badges? (Assume that no employees lose their badge a second time during the year.)

Subject: Session 6 Exercises – Answers

Preparer

Q.1

x	0	1	2	3	4	5	6	7	8	9	10	11	12
f(x)	.04	.055	.05	.045	.040	.035	.04	.045	0.05	.055	.06	.065	.07

x	13	14	15	16	17	18	19	20
f(x)	.08	.07	.06	.05	.04	.03	.02	0

(a) The probability of exactly 7 breakdowns in a 24-hour period, from the above table, is $p(7) = 0.045$

(b) The probability of more than 2 or less than 10 breakdowns is the probability of 3 through 9 breakdowns:

$$p_{3,9} = .045 + .04 + .035 + .04 + .045 + .05 + .055$$

$$= 0.31$$

(c) $E(x) = 0 \times .04 + 1 \times .055 + 2 \times .05 + \dots + 19 \times .02$

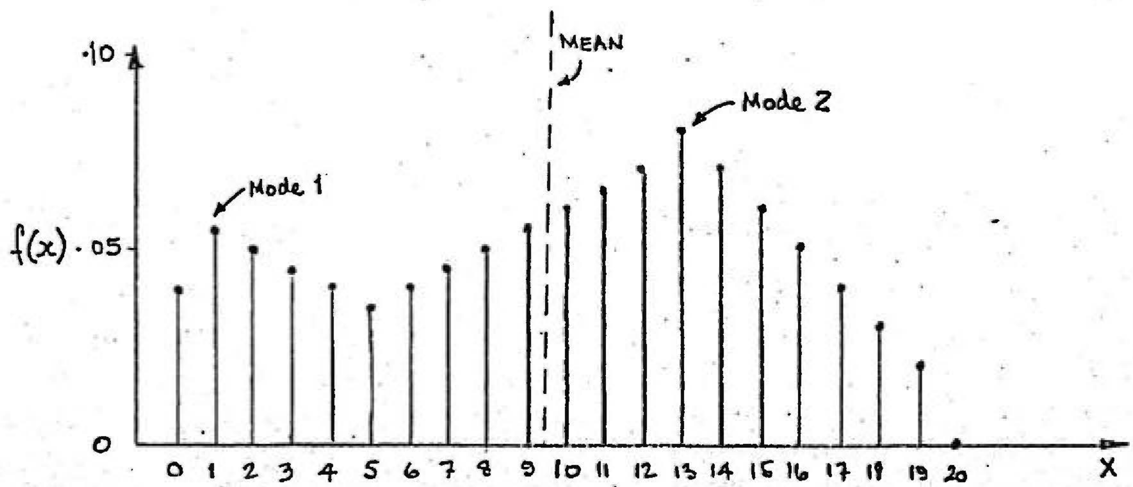
$= 9.55$

$V(x) = E(x^2) - \bar{x}^2 = 0 \times .04 + 1 \times .055 + (2)^2 \times .05 + (3)^2 \times .045 + \dots + (19)^2 \times .02 - (9.55)^2$

$= 119.46 - (9.55)^2$

$= 28.2575$

There are two modes, with values $x = 1$ and $x = 13$



Q.2

(a) The cumulative density function is obtained by integrating $f(x)$ over each of the ranges of x. The maximum value is given by:

$$F(x) = \int_0^3 0.1\,dx + \int_3^6 (a+bx)\,dx + \int_6^9 (c-dx)\,dx + \int_9^{12} 0.01\,dx$$

Performing the integration yields:

$$F(x) = \Big[0.1x\Big]_0^3 + \Big[ax + \frac{bx^2}{2}\Big]_3^6 + \Big[cx - \frac{dx^2}{2}\Big]_6^9 + \Big[0.01x\Big]_9^{12}$$

Using the values given for a, b, c, and d, this becomes

$$F(x) = \Big[0.1\times 3 - 0.1\times 0\Big] + \Big[\Big(0.05\times 6 + \frac{0.01}{2}\times 36\Big) - \Big(0.05\times 3 + \frac{0.01}{2}\times 9\Big)\Big]$$

$$+ \Big[\Big(0.1358\times 9 - \frac{0.001}{2}\times 81\Big) - \Big(0.1358\times 6 - \frac{0.001}{2}\times 36\Big)\Big] + \Big[0.01\times 12 - 0.01\cdot$$

$$= \Big[0.3\Big] + \Big[(0.48) - (0.195)\Big] + \Big[(1.1817) - (0.7968)\Big] + \Big[0.12 - 0.09\Big]$$

$$= 0.9999 \doteqdot 1.0$$

$$E(x) = 0.45 + 0.9 + 0.72 - 0.225 - 0.09 + 5.4999 - 0.243 - 2.4444$$
$$+ 0.072 + 0.72 - 0.405$$
$$= 4.9545$$

The expected length of time an employee will keep his or her badge is 4 months and 29 days.

(d) The variance is $E(X^2)$ and is defined by:

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx$$

Using the expression from computing $E(x)$, this is

$$E(x^2) = \int_0^3 0.1x^2\,dx + \int_3^6 (ax^2 + bx^3)\,dx + \int_6^9 (cx^2 - dx^3)\,dx + \int_9^{12} 0.01x^2\,d$$

$$= \left[\frac{0.1x^3}{3}\right]_0^3 + \left[\frac{0.05x^3}{3} + \frac{0.01x^4}{4}\right]_3^6 + \left[\frac{0.1358x^3}{3} - \frac{0.001x^4}{4}\right]_6^9$$
$$+ \left[\frac{0.01x^3}{3}\right]_9^{12}$$

$$= \left(\frac{0.1\times 27}{3}\right) + \left(\frac{0.05\times 216}{3} + \frac{0.01\times 1296}{4}\right] - \left(\frac{0.05\times 27}{3} + \frac{0.01\times 81}{4}\right)$$

$$+ \left(\frac{0.1358\times 729}{3} - \frac{0.001\times 6561}{4}\right) - \left(\frac{0.1358\times 216}{3} - \frac{0.001\times 1296}{4}\right)$$

$$+ \left(\frac{0.01\times 1728}{3}\right) - \left(\frac{0.01\times 729}{3}\right)$$

$$= 0.9 + 3.6 + 3.24 - 0.45 - 0.2025 + 32.9994 - 1.64025 - 9.777$$
$$+ 0.324 + 5.76 - 2.43$$

$$E(x^2) = 32.32305$$

Therefore, the standard deviation is:

$$\text{s.d.} = \sqrt{E(x^2)} = 5.685$$

(e) The expected cost is obtained by determining

$$E(C) = \int_{-\infty}^{\infty} f(x)g(x)\,dx$$

$$= \int_0^3 (.5 + .1(8+5x))\,0.1\,dx + \int_3^6 (.5 + .1(8+5x))(a+bx)\,dx$$

$$+ \int_6^9 (.5 + .1(8+5x))(c-dx)\,dx + \int_9^{12} (.5 + .1(8+5x))\,0.01\,dx$$

$$= \int_0^3 (.05 + .01(8+5x))\,dx + \int_3^6 (.5(a+bx) + .1(8+5x)(a+bx))\,dx$$

$$+ \int_6^9 (.5(c-dx) + .1(8+5x)(c-dx))\,dx + \int_9^{12} (.005 + .001(8+5x))\,dx$$

$$= \int_0^3 (.13 + .05x)\,dx + \int_3^6 (.065 + .11x + .005x^2)\,dx$$

$$+ \int_6^9 (.17654 + .0621x - .0005x^2)\,dx + \int_9^{12} (.013 + .005x)\,dx$$

Evaluating these integrals gives:

$$E(C) = \left[.13x + \frac{.05x^2}{2}\right]_0^3 + \left[.065x + \frac{.11x^2}{2} + \frac{.005x^3}{3}\right]_3^6 + \left[.17654x\right.$$

$$\left. + \frac{.0641x^2}{2} - \frac{.0005x^3}{3}\right]_6^9 + \left[.013x + \frac{.005x^2}{2}\right]_9^{12}$$

Subject Session 6 Exercises – Answers Preparer

$$E(C) = \left(.13\times3 + \frac{.05\times9}{2}\right) + \left(.065\times6 + \frac{.11\times36}{2} + \frac{.005\times216}{3}\right) - \left(.065\times3 + \frac{.11\times9}{2} + \frac{.005\times27}{3}\right) + \left(.17654\times9 + \frac{.0621\times81}{2} - \frac{.0005\times729}{3}\right) - \left(.17654\times6 + \frac{.0621\times36}{2} - \frac{.0005\times216}{3}\right) + \left(0.03\times12 + \frac{.005\times144}{2}\right) - \left(.013\times9 + \frac{.005\times81}{2}\right)$$

$$E(C) = 0.39 + 0.225 + 0.39 + 1.98 + 0.36 - 0.195 - 0.495 - 0.045 + 1.58886 + 2.51505 - 0.1215 - 1.05924 - 1.1178 + 0.036 + 0.156 + 0.36 - 0.117 - 0.2025$$

$$= \$4.65$$

The expected annual cost is $8,500 \times \$4.65 = \$39,506.90$

(f) The number of employees who have not lost their badges by 11.5 months is obtained by finding first the probabilit that an employee loses his or her badge in 11.5 to 12 month

$$f'(11.5) = \int_{11.5}^{12} 0.01x\,dx = \left[\frac{0.01x^2}{2}\right]_{11.5}^{12} = \left(\frac{0.01\times144}{2}\right) - \left(\frac{0.01\times1}{2}\right.$$

$$= \frac{11.75\times0.01}{2} = 0.05875$$

If there are 8,500 employees, then $0.05875 \times 8,500$

= 499 employees will not have lost their badges within 11.5 months.

8,001 employees will be charged the $5 fee for a total fee of $40,005. The annual replacement cost is $39,506.90. Hence, RTD would make a profit of $498.10.

SESSION 7 EXERCISES

1. You are conducting an on-board survey on RTD buses. Suppose that there are 40 people riding a bus that is boarded by a surveyor. Of the 40 people riding the bus, 18 paid a cash fare and 22 used a pass. A sample of 10 passengers is to be drawn randomly. What is the probability that the sample will include at least 5 pass users if:

(a) Sampling is conducted with replacement?

(b) Sampling is conducted without replacement?

(c) How would you design a nonprobability sample of bus riders?

2. In Session 2, question 2, you chose five bus lines from the Table 24. Estimate the mean and standard deviation of the cars per household for your sample and compare to the answers you obtained for mean and standard deviation in Question 2 of Session 3. What do you deduce from this?

Use the random numbers at the back of the text book to draw a random sample of five bus lines from Table 24. Again compute the mean and standard deviation. How does it compare to the values you just calculated?

3. Suppose that bus arrivals at Fifth and Hill is a Poisson process, such that, in a 5-minute period during the afternoon peak, lambda has a value (in the Poisson distribution) of 2.8.

(a) What is the probability that there will be exactly 3 buses passing through the intersection in a 5-minute period?

(b) What is the probability that there will be not more than 3 buses in the period?

(c) What is the probability that there will be more than 3 buses in the period?

4. If 52% of the voters in Los Angeles County favor a Proposition to add a further half-cent to the Sales Tax for transit in 1985, what is the probability that, in a sample of 50 voters interviewed at random, the majority will favor the Proposition?

If you were to design such a sample, would you sample with or without replacement, and why?

Q.1

(a) In sampling with replacement, the population becomes effectively infinite, so that a Bernoulli process can be assumed. Categorizing holding a pass as a "success," the probability of finding a success is 22/40 = 0.55. This is π in the Binomial Distribution.

We need to determine $P(R \geq 5 \mid n = 10, \pi = 0.55)$.

Because we need to use the cumulative table of the Binomial distribution, it is easier to reverse the definition of a success. Then the problem becomes "What is the probability that the sample will contain 5 or fewer cash-fare riders?"

This is defined by $P'(R' \leq 5 \mid n=10, \pi = 0.45)$. In the statistics book, no entries are provided for $\pi = 0.45$. To calculate this, evaluate:

$$P'(R' \leq 5 \mid n=10, \pi = 0.45) = \sum_{t=0}^{5} \binom{10}{t} (.45)^t (.55)^{10-t}$$

For $t=0$ $\quad \dfrac{10!}{0!\,10!} (.45)^0 (.55)^{10} = 0.0025$

$t=1$ $\quad \dfrac{10!}{1!\,9!} (.45)^1 (.55)^9 = 0.0207$

$t=2$ $\quad \dfrac{10!}{2!\,8!} (.45)^2 (.55)^8 = 0.0763$

$t=3$ $\quad \dfrac{10!}{3!\,7!} (.45)^3 (.55)^7 = 0.1665$

$t=4$

$\dfrac{10!}{4!\,6!} (.45)^4 (.55)^6 = 0.2384$

$t=5$

$\dfrac{10!}{5!\,5!} (.45)^5 (.55)^5 = 0.2340$

Hence $\sum_{t=0}^{5} \binom{10}{t} (0.45)^t (0.55)^{10-t} = 0.7384$ which is the desired probability.

(b) In sampling without replacement, the population is finite and the hypogeometric function must be used. To determine $P'(R' \leq 5 \mid n=10, N=40, D=18)$, we need to evaluate $\sum_{t=0}^{5} \frac{\binom{18}{t}\binom{22}{10-t}}{\binom{40}{10}}$. Again, we have to evaluate these manually:

$t=0 \quad \dfrac{\frac{18!}{0!\,18!} \times \frac{22!}{10!\,12!}}{\frac{40!}{10!\,30!}} = 0.0008$

$t=1 \quad \dfrac{\frac{18!}{1!\,17!} \times \frac{22!}{9!\,13!}}{\frac{40!}{10!\,30!}} = \dfrac{8,953,560}{847,660,528} = 0.0106$

$t=2 \quad \dfrac{\frac{18!}{2!\,16!} \times \frac{22!}{8!\,14!}}{847,660,528} = \dfrac{153 \times 319,770}{847,660,528} = 0.0577$

$t=3 \quad \dfrac{\frac{18!}{3!\,15!} \times \frac{22!}{7!\,15!}}{847,660,528} = \dfrac{816 \times 170,554}{847,660,528} = 0.1642$

$t=4 \quad \dfrac{\frac{18!}{4!\,14!} \times \frac{22!}{6!\,16!}}{847,660,528} = \dfrac{3060 \times 74,613}{847,660,528} = 0.2693$

$t=5 \quad \dfrac{\frac{18!}{5!\,13!} \times \frac{22!}{5!\,17!}}{847,660,528} = \dfrac{8568 \times 26,334}{847,660,528} = 0.2662$

Hence $\sum_{t=0}^{5} \frac{\binom{18}{t}\binom{22}{10-t}}{\binom{40}{10}} = 0.7688$ which is the desired probability.

Note that, in sampling without replacement, the probability that at least 5 pass holders will be found in the sample is higher than in sampling with replacement.

(c) While a number of nonprobability samples could be designed, the most obvious one is a systematic sample of every fourth boarding passenger.

Q.2

The sample I drew was a systematic sample of every tenth bus line. For my sample:

Mean cars per household = 1.507

Standard deviation = 1.289

In question 2, session 3, the mean was found to be 1.21 and the standard deviation was 1.196. Both values in the sample differ from the total population values and one would clearly err if one used only the 5-line sample.

To draw a random sample, assume first that the bus lines are renumbered from 1 through 45 (omitting those with no car ownership data). Using the random numbers from page 523, Starting on line 12 and going row (line) by row:

.15360 (x45) = 7 ; .86221 (x45) = 39 ; .06240 = 3 ; .68606 = 31;

.05993 = 3 reject already chosen ; .28257 = 13

This sample (without replacement) gives:

Line	None	One	Two	Three	Four	Five	Total
73	11	20	17	6	2	4	60
844	35	47	45	18	5	5	155
29	55	32	8	13	2	1	111
454	18	19	21	6	-	2	66
114	46	74	57	25	9	11	222
Total	165	192	148	68	18	23	614

Mean car ownership = 1.432

Standard deviation = 1.277

While both mean and standard deviation are less than for the previous sample, they are still larger than for the entire table. Some samples should generate values smaller than the table values, also. Again, it is clear that acceptance of the sample values as true and correct estimates of the table would be in error.

Q.3

The Poisson distribution is defined for a 5-minute period, with $\lambda = 3$. Each of the three questions deals with $t = 1$.

(a) The answer is obtained by evaluating the Poisson Mass Function for $r = 3$, $\lambda = 3$, $t = 1$
Using the Table G on pp. 562-563 for $\lambda t = 3$, $r = 3$
$f_p(3|\lambda, t) = 0.2240$ which is the desired probability.

(b) In this case, we use the cumulative Poisson terms from Table H, with $r = 3$.
$F(3) = 0.6472$ which is the desired probability

(c) The probability that there will be more than 3 buses in one 5-minute period is simply $(1 - F(3)) = 0.3528$. Note that the table shows that not more than 12 buses have a defined (non-zero) probability of passing the intersection in a 5-minute period.

Q.4

The voters in Los Angeles County can be considered an

infinite population, with successes and failures defined i
terms of favoring or not favoring the Proposition. Becau
we want to know the probability that more than 25
in the sample favor the proposition, estimate the probab
from a Binomial distribution that less than 25 voters
vote against the proposition:

$$P(R \leq 24 \mid n = 50, \pi = 0.48)$$

Using Tables E and F:

For $\pi = 0.40$, and summing from Table E

$$P(R \leq 24 \mid n = 50, \pi = 0.40) = 0.9022$$

For $\pi = 0.50$, from Table F

$$P(R \leq 24 \mid n = 50, \pi = 0.50) = 0.4439$$

Note that our computation for Q.1 (a) showed the value for 0.45 not to be halfway between 0.4 and 0.5 for that cas (P for 0.40 = 0.8337, P for 0.45 = 0.7384, P for 0.5 = .623c but to be slightly higher than the midpoint (which is P = 0.7 Nevertheless, a close approximation for the present problem be to interpolate linearly between 0.40 and 0.50.

Hence $P(R \leq 24 \mid n = 50, \pi = 0.48) \doteq 0.5356$

Therefore, the probability is approximately 0.54 that a majority of voters will favor the proposition from the sample of 50.

I would sample without replacement because:

(a) I will be asking individuals and no individual will like being asked more than once

(b) Because the population is effectively infinite, I do not need to use replacement to create a Binomial distribution, as would be the case for a finite population

(c) I desire to maximize the variance in the sample relative to the population variance, so that without replacement is the indicated method.

SESSION 8 EXERCISES

1. Define the action space, the parameter space, and the state space for the problem in Session 1, Exercise 1. Construct a testable hypothesis for the problem, including specifying the null hypothesis. Define what would be Type I error and what would be Type II error in this problem. For the hypothesis you construct, what would be your decision rule? (Hint: use one of the distributions discussed in Session 7.)

2. The RTD is considering mounting an advertizing campaign to attract more riders. The recommendation of an advertizing agency, retained to develop the campaign, is the claim that, even in the peak periods, at least 75% of RTD buses are on schedule. Before adopting this campaign, the RTD Board has instructed the General Manager to make sure that the claim is not untrue. You are directed by the General Manager to obtain the necessary information and determine whether or not the Board should approve the campaign.

(a) Identify the action space and state space of this problem.

(b) How would you partition the state space?

(c) State what your null and alternative hypotheses would be.

(d) What is the appropriate test statistic?

(e) What is the appropriate sampling distribution for the problem?

(f) Specify the domain of the sampling distribution.

(g) If the General Manager has adopted the .05 level of significance, what are the acceptance and rejection regions of the domain of the sampling distribution?

(h) What is the decision rule?

(i) Define the Type I and Type II errors in this problem.

Q.1

The Action Space is two elements:
(i) maintain fares at present 5(
(ii) raise fares to \$1.00 base

The Parameter Space is the vote on the referendum, with population values of interest being ≤ 5($> 50\%$.

The State Space is the same thing as the parameter sp

The appropriate testable hypothesis is:

H_0: Vote for the referendum $\leq 50\%$ of all votes

H_1: Vote for the referendum $> 50\%$ of all vo

Type I error would result from adopting the position t referendum will pass when it does not.

Type II error would result from adopting the positio the referendum will not pass when it does.

Assuming a level of significance of .05, the decision could be written as:

$$P(R \geq c \mid H_0 \text{ is true}) \leq .05$$

The Binomial distribution is an appropriate one for d the results of voting from a large population. In this case, a sample size must be selected and a limiting value must be selected for Ho. Using 0.50 as the limiting value, and using a sample size of 75 people:

$$P(R \geq c \mid n = 75, \pi = .50) \leq .05$$

c is the number of people who would have to indicate favoring the referendum. In order to be able to use the cumulative Binomial distribution, we need to consider a redefinition in terms of values of R less than or equal to a limiting value.

$$P(R \geq c \mid n = 75, \pi = .50) = 1 - P(R < c \mid n = 75, \pi = 0.50) \leq .05$$

$$P(R < c \mid n = 75, \pi = 0.50) = F(c-1 \mid n = 75, \pi = 0.50)$$ because c must be an integer value

Also $1 - .50 \leq P(R < c \mid n = 75, \pi = 0.50)$

Therefore $F(c-1 \mid n = 75, \pi = 0.50) \geq .95$

Using Table F from the book

$$F(45) = 0.9680, \text{ while } F(44) = 0.9473$$

Therefore $c - 1 = 45$, so $c = 46$.

The null hypothesis should be rejected if the number of people in a sample of 75 who indicate they would vote for the referendum is 46 or more.

Q.2

(a) The action space consists of two alternatives:

(i) Adopt the ad campaign

(ii) Reject the ad campaign (or modify the on-time claim)

The state space consists of all possible proportions of peak-buses that are not running late: $\{\pi \mid 0 \leq \pi \leq 1\}$

(b) The partitions of the state space would be:

$\{\pi \mid 0 \leq \pi < .75\}$ and $\{\pi \mid .75 \leq \pi \leq 1\}$

(c) The hypotheses should be:

$H_0: \ \pi < 0.75$

$H_1: \ \pi \geq 0.75$

(d) The appropriate test statistic is R, the number of buses in the sample that operate on time in the peak period

(e) The sampling distribution should be the Binomial distribution, with late arrival being defined as a "failure."

(f) To define the domain, a sample size must be selected. If the sample is 100 bus trips from the peak period, then the domain is:

$\{r \mid r = 0, 1, 2, 3, 4, \ldots, 100\}$

(g) The acceptance and rejection regions must be calculated from the Binomial distribution.

$$P(R > c \mid n = 100, \pi = 0.75) \leq .05$$

Rewriting this gives

$$P(R \leq c \mid n = 100, \pi = 0.75) \geq .95$$

$$F(c \mid n = 100, \pi = 0.75) \geq .95$$

From Table F with $n = 100$, $\pi = 0.75$

$F(c) = .9370$ for $r = 81$

$F(c) = .9624$ for $r = 82$

Therefore $c = 82$

The acceptance region is: $\{r \mid r = 0, 1, 2, 3, \ldots, 82\}$

The rejection region is: $\{r \mid r = 83, 84, 85, \ldots, 100\}$

(h) The decision rule: if $R \geq 83$, reject H_0 and proceed with the advertizing campaign; otherwise accept H_0 and do not proceed with the advertizing campaign.

(i) The Type I error would be to proceed with the ad campaign when less than 75 percent of RTD peak-period buses are on time.

Type II error would be to reject the advertizing campaign

RTD

Working Notes

Subtask | Date | Page 5 of 5

Subject: Session 8 Exercises – Answers | Preparer

when 75% or more of the buses do run on time.

SESSION 9 EXERCISES

1. Consider Table 24, as used in Session 2, Exercise 2. Repeat the random sampling of five bus lines that you performed in Session 7, Exercise 2, to produce 4 more samples. For each sample, calculate the mean and the standard deviation of car ownership. Sample without replacement for this exercise. Present the results in a tabular distribution, to see the effect of sampling for the mean and the standard deviation. Comment on the result.

Repeat the exercise now with sampling with replacement. What differences do you find in the results?

2. Define the sampling distribution of the sample mean for the on-board survey of Session 7, Exercise 1, under condition (b) (sampling without replacement). Using the distribution you have defined, determine the expected value and the variance of the sample mean.

3. An on-board survey on RTD buses has questioned the ages of bus riders. Current ridership on RTD is about 1.5 million rides per day, and probably represents about 500,000 or more individual people. From the survey, the mean age of a bus rider is determined to be 36 years, and the standard deviation is 17.65.

(a) If it was reasonable to assume that ages of riders are normally distributed, how many riders will be between 25 and 45 years old?

(b) Under the normal distribution assumption, how many are over 65 years old?

(c) How reasonable is it to assume a normal distribution? Explain your answer.

RTD Working Notes

Subtask: Date: Page 1 of 8

Subject: Session 9 Exercises – Answers

Preparer:

Q.1

My sample in Session 7 was:

Lines: 73, 844, 29, 454, and 114

Continuing the same sequence of random numbers produces:

.80451 (x45) = 36; .90422 = 41; .20624 = 9; .31777 = 14; .53092 = 24;

.10820 = 5; .71841 = 32; .58926 = 26; .48072 = 22; [.70936 = 32;] .33884 = 15;

.63004 = 28; [.81011 = 36;] [.90241 = 41;] .78293 = 35; .43707 = 20; .65260 = 29

.46559 = 21; .21593 = 10; [.61172 = 28;] [.05802 = 3;] [.48592 = 22;] .96801 = 44;

.88425 = 40; .24698 = 11; [.22898 = 10;] .26339 = 12

In Table 24, this produces the samples shown in Table 9-1, p. 2

Means and standard deviations are repeated below:

Sample	Mean	Standard Deviation
1	1.432	1.277
2	1.242	1.155
3	1.400	1.205
4	1.565	1.262
5	1.186	1.206
Total Table	1.210	1.196

(i) Most of the means are higher than that of the entire table, and so are most of the standard deviations. There are fairly wide variations in the mean – 1.186 to 1.565. – but much less variation in the

Subtask | Date | Page 2 of 8

Subject Session 9 Exercises – Answers | Preparer

TABLE 9-1

Bus Line	None	One	Two	Three	Four	Five or More	Total
73	11	20	17	6	2	4	60
844	35	47	45	18	5	5	155
29	55	32	8	13	2	1	111
454	18	19	21	6	–	2	66
114	46	74	57	25	9	11	222
Total	165	192	148	68	18	23	614
826	28	28	13	5	1	2	77
861	36	46	29	18	2	4	135
86	47	64	34	11	6	-	162
152	26	31	31	8	8	3	107
210	72	71	49	13	4	3	212
Total	209	240	156	55	21	12	693
44	94	77	50	16	2	3	242
484	19	17	13	5	1	1	56
424	22	19	27	10	4	4	86
169	36	53	51	19	8	5	172
156	14	32	47	24	4	3	124
Total	185	198	188	74	19	16	680
431	26	27	30	10	6	2	101
822	19	20	14	4	2	4	63
166	20	25	15	12	3	2	77
435	33	51	42	29	3	4	162
168	5	17	21	9	4	3	59
Total	103	140	122	64	18	15	462
88	26	31	23	11	3	2	96
871	46	71	46	17	5	7	192
846	60	55	57	28	6	8	214
89	121	69	28	4	6	4	232
91	55	59	26	8	3	1	152
Total	308	285	180	68	23	22	886

standard deviations – 1.155 to 1.277.

(ii) The mean of the means is 1.365 and the mean of the standard deviations is 1.221. The mean of the means is, as expected, higher than the table mean, while the mean of the standard deviations is quite close to the standard deviation.

(iii) More correctly, comparison should be made to the 45 lines considered for sampling. In this case, the mean is 1.388 and the standard deviation is 1.245. Against these values, we see that 3 means are higher, 2 are lower and the mean of the means is lower; 2 standard deviations are higher, 3 are lower and the mean standard deviation is lower.

Sampling with Replacement

The samples now are: 73, 844, 29, 454, 29
114, 826, 861, 86, 152
210, 44, 484, 424, 169
484, 156, 431, 826, 861
822, 166, 435, 168, 88

For these samples, the values are given on p. 4

Sample	0	1	2	3	4	5+	Total
1	174	150	99	56	11	13	503
2	183	243	154	67	26	20	693
3	243	237	190	63	19	16	768
4	123	150	132	62	14	12	493
5	103	144	115	65	15	15	457

	Mean	Standard Deviation
1	1.243	1.241
2	1.380	1.236
3	1.253	1.176
4	1.482	1.209
5	1.540	1.255
Mean	1.374	1.223

(i) The mean of the means is higher than for the sample replacement and is closer to the Table mean. The mean of the standard deviations is almost the same. The range of the is smaller (as should be expected), as is the range of the standard deviations (as expected).

While the samples are small, the expected differences, of variation in the sample <u>with</u> replacement and results clo to the table mean, are borne out by the results.

Q.2

10 passengers are to be sampled without replacement from a population with 18 fare payers and 22 pass users. The mean of interest is the mean cash fare paid.

There are eleven outcomes of interest: 0 cash fares, 1 cash fare, 2 cash fares, ..., 10 cash fares. The frequency of these can be determined from the hypogeometric distribution. Six of the relative frequencies were determined in Session 7, Q.1b. Evaluation for t = 6 through 10 is shown on p.6.

Sample	$\bar{X}$	Relative Frequency
0 cash fares	0	.0008
1 cash fare	\$ .05	.0106
2 cash fares	\$ .10	.0577
3 cash fares	\$ .15	.1642
4 cash fares	\$ .20	.2693
5 cash fares	\$ .25	.2662
6 "	\$ .30	.1602
7 "	\$ .35	.0578
8 "	\$ .40	.0119
9 "	\$ .45	.0013
10 "	\$ .50	.0001

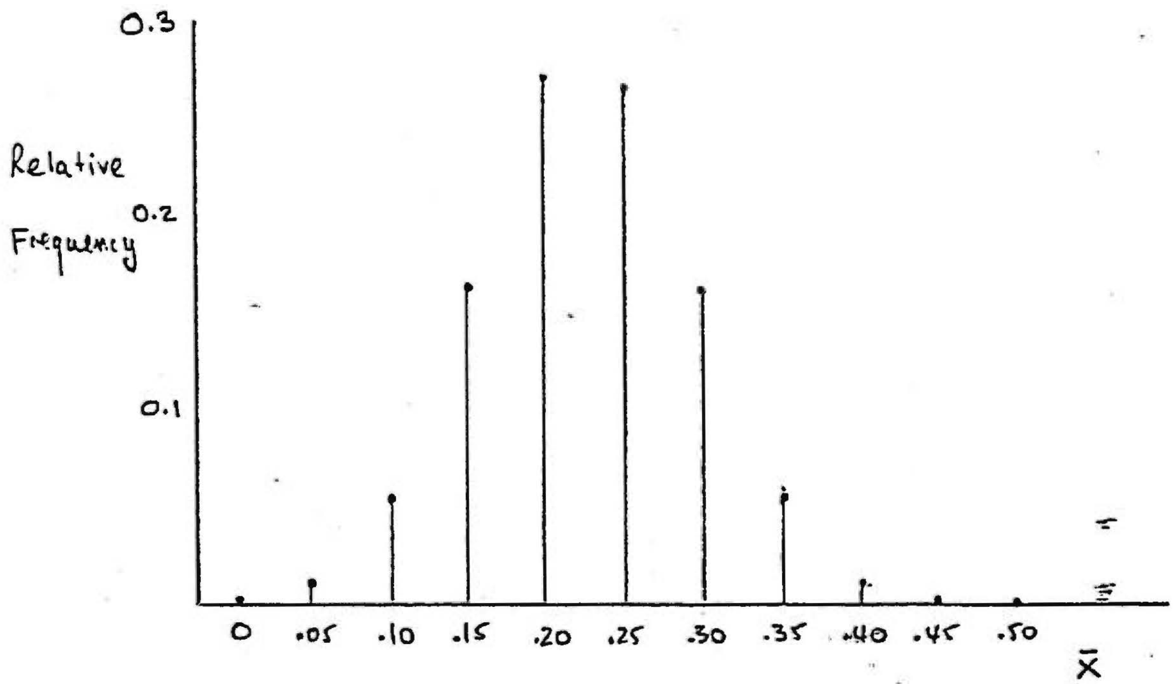
$t = 6$ $$\frac{\dfrac{18!}{6!\,12!} \times \dfrac{22!}{4!\,18!}}{847,660,528} = \frac{18564 \times 7315}{847,660,528} = 0.1602$$

$t = 7$ $$\frac{\dfrac{18!}{7!\,11!} \times \dfrac{22!}{3!\,19!}}{847,660,528} = \frac{31824 \times 1540}{847,660,528} = 0.0578$$

$t = 8$ $$\frac{\dfrac{18!}{8!\,10!} \times \dfrac{22!}{2!\,20!}}{847,660,528} = \frac{43758 \times 231}{847,660,528} = 0.0119$$

$t = 9$ $$\frac{\dfrac{18!}{9!\,9!} \times \dfrac{22!}{1!\,21!}}{847,660,528} = \frac{48620 \times 22}{847,660,528} = 0.0013$$

$t = 10$ $$\frac{\dfrac{18!}{10!\,8!} \times \dfrac{22!}{0!\,22!}}{847,660,528} = \frac{43758}{847,660,528} = 0.0001$$



Note that the distribution is very close to a bell-shaped normal curve.

$$E(\bar{x}) = 0 \times .0008 + .05 \times .0106 + .10 \times .0577$$

$$+ \ldots + .50 \times .0001 = 0.225$$

$$= 22.5 \text{ cents}$$

$$V(\bar{x}) = E(x^2) - (E(\bar{x}))^2 = 0.055135 - 0.$$

$$= 0.00451$$

The standard deviation of $\bar{x}$ = 0.06716

Q.3

(a) To use the standard normal distributi

$$Z = \frac{X - \bar{x}}{\sigma} = \frac{X - 36}{17.65}$$

This allows translation of $X = 25$ and $x =$

values of Z:

$$Z_1 = \frac{25 - 36}{17.65} = -0.623$$

$$Z_2 = \frac{45 - 36}{17.65} = 0.510$$

Using Table I from the text book:

for $Z_1 = -0.623$ $F_N(Z_1) = 0.2676$

for $Z_2 = 0.510$ $F_N(Z_2) = 0.6950$

The probability of obtaining a value of betw

45 is given by $F_N(Z_2) - F_N(Z_1) = 0.$

Hence, the proportion of riders with ages between 25 and 45 is 0.4274, or 42.74 percent which is 213,700 approximately.

(b) To find the proportion over 65:

$$Z_3 = \frac{65-36}{17.65} = 1.643$$

$F_N(Z_3) = 0.9495$ which is the proportion of riders under 65 years old. Therefore, the proportion over 65 is $1-0.9495 = 0.0505$. The approximate number of riders over 65 is, therefore 25,250.

(c) The assumption is <u>not</u> a very reasonable one. The normal distribution ranges from $-\infty$ to $+\infty$, while ages range from 0 to somewhere around 100. The distribution of bus rider ages is likely to be bimodal, with peaks for riders of school age and for the elderly. Within 3 standard deviations of the mean (which in the normal curve represents the largest portion of the population), the age range would be -16.95 to 88.95, which is not appropriate. Also, the median and mode of rider age would be 36 for the normal distribution.

SESSION 10 EXERCISES

1. In the Los Angeles region, the census estimates that average car ownership per household is 1.95 cars per household. What test would you perform to determine whether or not the car ownership of bus riders, as reported in Table 24 (Session 2, Exercise 2) is statistically different from regional car ownership? Perform the test, and comment on the result.

2. In Session 2, Exercises 1 and 2, two tables were provided -- Table 37 and Table 24. Table 37 contains responses from cash-fare-paying riders only, while Table 24 contains responses from all riders. From these two tables, one can compute the proportion of responding, cash-fare-paying riders as a proportion of all riders. Compute this proportion for each bus line in the Tables.

(a) How would you determine if there is statistical evidence that the proportions differ significantly from line to line? Perform the test.

(b) The mean proportion of responding, cash-fare-paying riders is 0.4185 (2302/5500), according to the tables. Is the (unweighted) mean of the proportions by bus line a closer estimate of this than the median? At a significance level of .95, is there any difference?

(c) Using your sampled bus lines from Table 24 (Session 9, Exercise 1), what is the expected mean and standard deviation of the number of respondents, based on these samples? Test whether or not these values are significantly different from the actual mean and standard deviation computed from the entire Table. From your samples, which of the sample means and standard deviations are significantly different from the Table mean and standard deviation, at a .95 level of significance?

Session 10 Exercises – Answers

Q.1

Because the population variance is not known, the appropriate test is the t-test.

For this test, we know:

Mean car ownership of all households = 1.95

Mean car ownership of bus riders = 1.21

Variance of car ownership = 1.4304

Variance of the sample mean $= \frac{S^2}{n} = \frac{1.4304}{5500}$

$= 0.00026$

Standard deviation of the sample mean

$S_{\bar{x}} = 0.01613$

Hence $t = \frac{1.21 - 1.95}{0.01613} = -45.88$

The table value of t for a two-tailed test with $n = \infty$ and at 95% confidence is −1.96

Therefore, the null hypothesis that the two means are identical should be rejected.

RTD

Working Notes

Session 10 Exercises – Answers

Q.2

Line #	Cash-Fare Riders	Total Riders	Proportion of Cash-Fare	Line #	Cash-Fare Riders	Total Riders	Proportion of Cash Fare
12	44	107	0.411	210	64	212	0.302
18	20	67	0.299	354	15	49	0.306
29	37	111	0.333	424	30	86	0.349
32	42	84	0.500	425	71	225	0.316
44	108	242	0.446	431	50	101	0.495
47	37	73	0.507	435	64	162	0.395
73	12	60	0.200	451	30	62	0.484
81	58	188	0.309	452	18	29	0.621
86	82	162	0.506	453	6	36	0.167
88	41	96	0.427	454	20	66	0.303
89	70	232	0.302	484	26	56	0.464
91	64	152	0.421	488	49	143	0.343
96	12	22	0.545	813	26	68	0.382
114	108	222	0.486	821	19	39	0.487
152	56	107	0.523	822	31	63	0.492
155	14	38	0.368	826	48	77	0.623
156	53	124	0.427	831	23	60	0.383
157	68	134	0.507	840	56	109	0.514
160	16	61	0.262	844	64	155	0.413
164	35	93	0.376	846	119	214	0.556
165	30	78	0.385	861	56	135	0.415
166	35	77	0.455	867	36	87	0.414
168	26	59	0.441	869	70	155	0.452
169	80	172	0.465	871	97	192	0.505
175	48	100	0.480	872	18	57	0.316

(a) The appropriate test is to take an unweighted average of the 50 lines and test this for a statistical difference from the mean of the table (which is a weighted mean). The true mean is 0.4185, and the mean of the 50 proportions is 0.41756. The standard deviation of the proportions is 0.0974. The variance is 0.0094869.

RTD 44-9
EFF 1/82

Therefore the estimated standard deviation
mean = $\sqrt{0.0094869/50}$ = 0.01377

The t statistic is given by:

$$t = \frac{0.41756 - 0.4185}{0.01377} = -0.06$$

This small value of t indicates that one ca[n]
null hypothesis that the unweighted mean i[s]
the overall mean.

(Note: there is no test for an individual val[ue]
the mean, because an individual value ha[s]
or standard deviation.)

(b) The mean of the bus line proportion is
median is obtained by rank ordering the l[ine]
proportions. This is done most easily by c[ounting]
lines that fall into certain groups:

.10–.199	.20–.299	.300–.399	.400–.499	.500–
I	III	~~IIII~~ ~~IIII~~ ~~IIII~~ I	~~IIII~~ ~~IIII~~ ~~IIII~~ IIII	~~IIII~~ I
1	3	16	19	9

The median is the 5th value in the .400–.49[9]
this is 0.421. Clearly, the mean of the li[ne]

Again, the t test is the appropriate test, using the median (which does not have a calculable variance) as the value to test against.

Thus: $$t = \frac{0.41756 - 0.421}{0.01377} = -0.250$$

This value is much less than the table value for .05 significance, so that the null hypothesis cannot be rejected at this confidence level.

(c) Mean respondents and standard deviations (using sample - expected - s.d.) are:

Sample #	Mean	S.d.
1	122.8	67.42922215
2	138.6	51.81988036
3	136.0	73.44385611
4	92.4	42.23505653
5	177.2	54.32494823

The expected mean from these five samples is simply $1/5 \sum_{i=1}^{5} x_i = 133.4$. The standard deviation of the means = 30.60065359.

A10P4

Session 10 Exercises – Answers

The average number of respondents per bus line
with a standard deviation of 58.365. We can es
the standard deviation of the Table mean from

$$s.d._{\bar{x}} = 8.3955$$

Because both standard deviations are known, a
can be performed:

$$Z = \frac{(\bar{x}_s - \bar{x})}{\sigma_{(\bar{x}_s - \bar{x})}} = \frac{133.4 - 110}{\sqrt{(30.60061^2 + 8.5955^2)}} = 0.$$

At 95% confidence for a two-tailed test, Z = 1.96.
the calculated Z is much less than this, the
hypothesis – that the two means are equal –
be rejected.

For the variances, the appropriate test is the F

$$F = \frac{S_1^2}{S_2^2}$$

S_1^2 = mean of the variances of the 5 samples

= 3472.2

S_2^2 = 3524.2196

Therefore $F = \frac{3472.2}{3524.2196} = 0.9852$

Degrees of freedom for this F test are 24 (25 minus 1) and 49. At 0.95 and these degrees of freedom, the table value of F is approximately 1.75.
Because the calculated value of F is less than the table value of F at 0.95, the null hypothesis cannot be rejected.

The largest numerically different mean is that of sample 5, for which:

$$\sigma_{\bar{x}} = 24.29485542$$

Therefore

$$Z = \frac{(177.2 - 110)}{\sqrt{(24.29485542^2 + 8.3955^2)}} = 2.614$$

Since $Z = 1.96$ at 0.95 for a two-tailed test, this sample mean is significantly different from the table mean.

The next largest difference is for sample 2 and sample 3, both of which are around 28.
For sample 2:

$$Z = \frac{(28.6)}{\sqrt{[(51.81988036)^2/5 + (8.3955)^2]}} = 1.160$$

This is not significantly different, so no other samples will be.

Similarly, for standard deviations, the largest difference is for sample 4, for which:

$$F = \frac{(42.23505653)^2}{3524.2196} = 0.506 \qquad \nu_1 = 4,\ \nu_2 = 49$$

The value of F is less than the table value of 2.57 at .95, and is greater than the value of .175 at .05. Therefore, these standard deviations are not significantly different at 95%.

INTRODUCTION TO PROBABILITY

AND STATISTICS

FINAL TAKE-HOME EXAMINATION

You are in charge of a project to assess the viability of ir the Harbor Freeway, to run from the Sepulveda Boulevard inte Street intersection, where buses will proceed onto downtown

a) It is intended to permit carpools over a certain size tc it is intended to run 120 buses per hour in the peak one hou Assume that the capacity of the busway in addition to the 12 vehicles per hour and that capacity must not be exceeded 95 Vehicle occupancy studies by CALTRANS show that carpools in by a binomial distribution, such that for carpools of 3 or n proportion of carpools is 0.15,and for 4 or more people it i that the peak volume into the city on the Harbor Freeway is hour, what size carpool can be allowed to share the busway?

b) A sample survey of 25 residents in the South Bay Area wh Los Angeles has been undertaken. Sampling was done using a without replacement. Among the questions asked are:

- How many cars are available to your household?
- Would you use an express bus on a Harbor Freeway bus work?
- How many times have you ridden a bus in the past mor

Data from the survey are provided in Table 1, attached.

1. What is the average number of cars available, and standard deviation of the number available?
2. What is the median number of times a respondent l the past month?
3. What proportion of respondents would ride the bus
4. Compute the mean, variance, and standard deviati times a respondent has used bus in the past montl
5. What is the probability that a respondent with a use the busway, and what is the probability that no car available will not use the busway?
6. What is the standard deviation of the mean of th available?

7. If the 1980 Census showed that mean car availability for households in the South Bay Area is 1.49, with what confidence could you state that the mean of the sample is the same as the South Bay mean?

8. What type of a standard distribution best describes the responses to the question on riding the busway? Based on this sample, what is the probability that, from a subsample of 10 persons, all 10 would say they would ride the busway? What is the probability that none of them would?

9. In a subsequent survey of 500 residents of South Bay, it is found that the mean number of cars available to residents who work downtown is 1.44 with s = 0.8784, that the mean bus use of the respondents is 4 times per month with standard deviation of 7.3348, and that 35 percent of respondents indicate they would be likely to ride the bus. Based on these subsequent numbers, you decide that you doubt that the original sample of 25 was a truly random sample. Structure the hypotheses you would seek to test, and set out the details of the tests you would make. What is your decision rule? Conduct the tests and draw your conclusions.

c) From data collected by the SCRTD, it is known that 40.0 percent of SCRTD riders use a pass. Assuming that pass use is a Bernoulli process, what is the probability that 25 riders on a busway bus with 50 riders boarding will be pass users? What is the probability that at least 20 riders will pay a cash fare?

TABLE 1

Respondent	Cars Available	Ride Busway	Times Used Bus in Past Month
1	0	Yes	15
2	2	Yes	4
3	1	No	0
4	1	Yes	20
5	3	No	12
6	0	Yes	8
7	2	Yes	16
8	2	No	3
9	1	No	0
10	0	Yes	18
11	3	No	2
12	4	Yes	2
13	1	Yes	20
14	0	No	26
15	2	Yes	18
16	2	No	6
17	0	No	2
18	1	Yes	14
19	1	Yes	12
20	1	Yes	6
21	1	Yes	16
22	0	Yes	22
23	3	No	0
24	1	Yes	20
25	1	No	8