

STAT 6800 Homework 1

Augustine Ennin

September 2025

Q1.(a)

```
1 DATA z;  
2 INPUT x y;  
3 DATALINES;  
4 1 3.2  
5 2 4.5  
6 3 5.2  
7 4 9.3  
8 5 11.6  
9 ;  
10 PROC PRINT DATA = z;
```

Obs	x	y
1	1	3.2
2	2	4.5
3	3	5.2
4	4	9.3
5	5	11.6

Figure 1: Data Output

Q1.(b)

```
1 PROC MEANS DATA = z MEAN STD;  
2 VAR x y;  
3 RUN;
```

The MEANS Procedure

Variable	Mean	Std Dev
x	3.0000000	1.5811388
y	6.7600000	3.5387851

Figure 2: Mean and Standard Deviation

From Figure 2, the mean and standard deviation of x and y are 3.0(1.58) and 6.76(3.54), respectively.

Q2.(a)

```
1 DATA china_trade;  
2 INFILE "/home/u63997979/sasuser.v94/Elliott and Morrell/China#1.dat.txt";  
3 INPUT year total exports imports;  
4 balance = exports - imports;  
5 RUN;  
6 PROC PRINT DATA = china_trade;  
7 TITLE "China Exports, Imports, and Trade Balance";  
8 RUN;
```

China Exports, Imports, and Trade Balance

Obs	year	total	exports	imports	balance
1	1955	3.15	1.41	1.73	-0.32
2	1956	3.21	1.65	1.56	0.09
3	1957	3.10	1.60	1.51	0.09
4	1958	3.87	1.98	1.89	0.09
5	1959	4.38	2.26	2.12	0.14
6	1960	3.81	1.86	1.95	-0.09
7	1961	2.94	1.49	1.45	0.04
8	1962	2.66	1.49	1.17	0.32
9	1963	2.92	1.65	1.27	0.38
10	1964	3.46	1.92	1.55	0.37
11	1965	4.25	2.23	2.02	0.21
12	1966	4.61	2.37	2.25	0.12
13	1967	4.16	2.14	2.02	0.12
14	1968	4.05	2.10	1.95	0.15
15	1969	4.03	2.20	1.83	0.37
16	1970	4.59	2.26	2.33	-0.07
17	1971	4.84	2.64	2.21	0.43
18	1972	6.30	3.44	2.86	0.58
19	1973	10.98	5.82	5.16	0.66
20	1974	14.57	6.95	7.62	-0.67
21	1975	14.75	7.26	7.49	-0.23
22	1976	13.43	6.86	6.58	0.28
23	1977	14.80	7.59	7.21	0.38
24	1978	20.64	9.75	10.89	-1.14
25	1979	29.33	13.66	15.68	-2.02
26	1980	37.82	18.27	19.55	-1.28
27	1981	40.38	20.89	19.48	1.41
28	1982	39.30	21.82	17.48	4.34
29	1983	40.73	22.20	18.53	3.67
30	1984	49.77	24.42	25.36	-0.94
31	1985	60.25	25.92	34.33	-8.41
32	1986	60.10	27.01	33.08	-6.07
33	1987	68.11	34.71	33.40	1.31
34	1988	80.49	40.64	39.85	0.79
35	1989	82.58	43.44	39.14	4.30

Figure 3: China Exports, Imports, and Trade Balance

Q2.(b)

```

1 PROC MEANS DATA = china_trade MEAN MIN MAX;
2 VAR exports imports balance;
3 RUN;
4
5 PROC SGPLOT DATA = china_trade;
6 SERIES X=year Y=exports / LINEATTRS=(COLOR=blue THICKNESS=2);
7 SERIES X=year Y=imports /LINEATTRS=(COLOR=red THICKNESS=2);
8 SERIES X=year Y=balance / LINEATTRS=(COLOR=green THICKNESS=2);
9 TITLE "China Exports, Imports, and Trade Balance Over Time";
10 XAXIS LABEL="YEAR";
11 YAXIS LABEL="BILLIONS OF USD";
12 RUN;

```

The MEANS Procedure

Variable	Mean	Minimum	Maximum
exports	10.6828571	1.4100000	43.4400000
imports	10.7000000	1.1700000	39.8500000
balance	-0.0171429	-8.4100000	4.3400000



Figure 4: Time Series Plot

Imports and exports in China ranged from about \$1.17 billion to \$39.85 billion and \$1.41 billion to \$43.44 billion, respectively. On average, China exported \$10.68 billion worth of goods and services, while imports averaged \$10.70 billion. From the 1950s through the early 1970s, both exports and imports remained relatively flat, each staying below \$5 billion. Beginning in the mid-1970s, trade activity increased, with exports and imports surpassing \$10 billion. During the late 1970s and especially throughout the 1980s, both grew rapidly, with values exceeding \$40 billion by the end of the decade. For much of the period before the mid-1970s, the trade balance remained close to zero. From the 1980s through 1990, the balance fluctuated, alternating between trade surpluses and deficits.

Q2.(c)

```

1 DATA china_trade_decade;
2 SET china_trade;
3 IF 1950 <= year < 1960 THEN decade = 1950;
4 ELSE IF 1960 <= year < 1970 THEN decade = 1960;
5 ELSE IF 1970 <= year < 1980 THEN decade = 1970;
6 ELSE IF 1980 <= year < 1990 THEN decade = 1980;
7 RUN;
8 PROC PRINT DATA = china_trade_decade;
9 BY decade;
10 VAR year total exports imports balance;
11 TITLE "China Exports, Imports, and Trade Balance by Decade";
12 RUN;

```

China Exports, Imports, and Trade Balance by Decade

decade=1950					
Obs	year	total	exports	imports	balance
1	1955	3.15	1.41	1.73	-0.32
2	1956	3.21	1.65	1.56	0.09
3	1957	3.10	1.60	1.51	0.09
4	1958	3.87	1.98	1.89	0.09
5	1959	4.38	2.26	2.12	0.14

decade=1960					
Obs	year	total	exports	imports	balance
6	1960	3.81	1.86	1.95	-0.09
7	1961	2.94	1.49	1.45	0.04
8	1962	2.66	1.49	1.17	0.32
9	1963	2.92	1.65	1.27	0.38
10	1964	3.46	1.92	1.55	0.37
11	1965	4.25	2.23	2.02	0.21
12	1966	4.61	2.37	2.25	0.12
13	1967	4.16	2.14	2.02	0.12
14	1968	4.05	2.10	1.95	0.15
15	1969	4.03	2.20	1.83	0.37

Figure 5: China Trade Data Grouped by Decade

decade=1970					
Obs	year	total	exports	imports	balance
16	1970	4.59	2.26	2.33	-0.07
17	1971	4.84	2.64	2.21	0.43
18	1972	6.30	3.44	2.86	0.58
19	1973	10.98	5.82	5.16	0.66
20	1974	14.57	6.95	7.62	-0.67
21	1975	14.75	7.26	7.49	-0.23
22	1976	13.43	6.86	6.58	0.28
23	1977	14.80	7.59	7.21	0.38
24	1978	20.64	9.75	10.89	-1.14
25	1979	29.33	13.66	15.68	-2.02

decade=1980					
Obs	year	total	exports	imports	balance
26	1980	37.82	18.27	19.55	-1.28
27	1981	40.38	20.89	19.48	1.41
28	1982	39.30	21.82	17.48	4.34
29	1983	40.73	22.20	18.53	3.67
30	1984	49.77	24.42	25.36	-0.94
31	1985	60.25	25.92	34.33	-8.41
32	1986	60.10	27.01	33.08	-6.07
33	1987	68.11	34.71	33.40	1.31
34	1988	80.49	40.64	39.85	0.79
35	1989	82.58	43.44	39.14	4.30

Figure 6: China Trade Data Grouped by Decade

Q2.(d)

```

1 PROC PRINT DATA = china_trade;
2 WHERE year >= 1980;
3 ID year;
4 VAR total exports imports balance;
5 TITLE "China Exports, Imports, and Balance (1980 and Later)";
6 RUN;

```

China Exports, Imports, and Balance (1980 and Later)

year	total	exports	imports	balance
1980	37.82	18.27	19.55	-1.28
1981	40.38	20.89	19.48	1.41
1982	39.30	21.82	17.48	4.34
1983	40.73	22.20	18.53	3.67
1984	49.77	24.42	25.36	-0.94
1985	60.25	25.92	34.33	-8.41
1986	60.10	27.01	33.08	-6.07
1987	68.11	34.71	33.40	1.31
1988	80.49	40.64	39.85	0.79
1989	82.58	43.44	39.14	4.30

Figure 7: China Trade Data for 1980 and Later

Q3.(a)

```

1 DATA BTT;
2 INFILE "/home/u63997979/sasuser.v94/Elliott and Morrell/btt.dat.txt";
3 INPUT
4 childid 1-4
5 sex 6
6 bweight 8-11
7 gestage 13-14
8 momage 16-17
9 parity 19
10 mdbp 21-23
11 msbp 25-27
12 momeduc 29
13 mmedaid 31
14 socio 33
15 dbp5 35-37
16 sbp5 39-41
17 ht5 43-47
18 wt5 49-52
19 hdl5 54-57
20 ldl5 59-62
21 trig5 64-67
22 smoke5 69
23 medaid 71
24 socio5 73;
25 bmi = wt5 / (ht5)**2;
26 RUN;
27 PROC PRINT DATA = BTT (OBS=10);
28 VAR childid ht5 wt5 bmi;
29 TITLE "BMI for the first 10 Observation"
30 RUN;

```

BMI for the first 10 Observation RUN

Obs	childid	ht5	wt5	bmi
1	1	1.025	13.5	12.8495
2	2	1.123	18.5	14.6694
3	3	1.090	19.8	16.6653
4	4	1.113	20.4	16.4680
5	5	1.065	18.3	16.1344
6	6	1.063	17.3	15.3102
7	7	1.112	19.2	15.5271
8	8	1.134	19.5	15.1638
9	9	1.039	16.6	15.3772
10	10	1.118	17.9	14.3209

Figure 8: BMI for the first 10 children

Q3.(b)

```

1 DATA childHealth;
2 SET BTT;
3 IF msbp >= 140 AND mdbp >= 90 THEN htn = 4;
4 ELSE IF msbp >= 140 AND mdbp < 90 THEN htn = 3;
5 ELSE IF msbp < 140 AND mdbp >= 90 THEN htn = 2;
6 ELSE IF msbp < 140 AND mdbp < 90 THEN htn = 1;
7 RUN;
8 PROC FORMAT;
9 VALUE htn_fmt
10 1 = "Normotensive"
11 2 = "Isolated Diastolic Hypertension"
12 3 = "Isolated systolic hypertension"
13 4 = "Hypertension";
14 PROC PRINT DATA = childhealth (OBS=30);
15 FORMAT htn htn_fmt.;
16 VAR childid msbp mdbp htn;
17 RUN;

```


Obs	childid	msbp	mdbp	htn
1	1	110	70	Normotensive
2	2	110	80	Normotensive
3	3	90	50	Normotensive
4	4	110	60	Normotensive
5	5	120	80	Normotensive
6	6	100	60	Normotensive
7	7	100	70	Normotensive
8	8	110	80	Normotensive
9	9	90	60	Normotensive
10	10	120	80	Normotensive
11	11	100	50	Normotensive
12	12	120	80	Normotensive
13	13	110	60	Normotensive
14	14	120	80	Normotensive
15	15	110	70	Normotensive
16	16	120	80	Normotensive
17	17	110	70	Normotensive
18	18	90	60	Normotensive
19	19	120	70	Normotensive
20	20	100	70	Normotensive
21	21	100	70	Normotensive
22	22	100	70	Normotensive
23	23	130	90	Isolated Diastolic Hypertension
24	24	80	60	Normotensive
25	25	110	80	Normotensive
26	26	120	80	Normotensive
27	27	110	80	Normotensive
28	28	120	70	Normotensive
29	29	110	70	Normotensive
30	30	100	60	Normotensive

Figure 9: Classification of the Mother's BP at child's birth

Q4.(a)

```

1 DATA One;
2 INPUT ID Name $ Prescore;
3 DATALINES;
4 11 Joe 19
5 13 Darcy 22
6 12 Ted 21
7 14 Jenny 23
8 16 Chris 17
9 18 Jane 21
10 19 Jeff 24
11 15 Bill 18
12 17 Bill 22
13 ;
14 RUN;
15 PROC PRINT DATA = One;

```

Obs	ID	Name	Prescore
1	11	Joe	19
2	13	Darcy	22
3	12	Ted	21
4	14	Jenny	23
5	16	Chris	17
6	18	Jane	21
7	19	Jeff	24
8	15	Bill	18
9	17	Bill	22

Figure 10: One

```

1 DATA Two;
2 INPUT ID Postscore;
3 DATALINES;
4 12 22
5 11 39
6 13 42
7 17 42
8 14 21
9 19 18
10 15 11
11 16 37
12 20 21
13 ;
14 RUN;
15 PROC PRINT DATA = Two;

```

Obs	ID	Postscore
1	12	22
2	11	39
3	13	42
4	17	42
5	14	21
6	19	18
7	15	11
8	16	37
9	20	21

Figure 11: Two

Q4.(b)

```

1 PROC SORT DATA = One; BY ID; RUN;
2 PROC SORT DATA = Two; BY ID; RUN;
3
4 DATA OneTwo;
5 MERGE One Two;
6 BY ID;
7 RUN;
8
9 PROC PRINT DATA = OneTwo;
10 TITLE "Merged Pre/Post Dataset: OneTwo";
11 RUN;
```

Merged Pre/Post Dataset: OneTwo

Obs	ID	Name	Prescore	Postscore
1	11	Joe	19	39
2	12	Ted	21	22
3	13	Darcy	22	42
4	14	Jenny	23	21
5	15	Bill	18	11
6	16	Chris	17	37
7	17	Bill	22	42
8	18	Jane	21	.
9	19	Jeff	24	18
10	20		.	21

Figure 12: OneTwo

It is observed that ID 18 has a missing value for the Postscore variable, while ID 20 has no entry for Name and also lacks a value for the Prescore variable.

Q4.(c)

```
1 DATA OneTwo2;  
2 SET OneTwo;  
3 IF Prescore = . OR Postscore = . THEN DELETE;  
4 RUN;  
5 PROC PRINT DATA = OneTwo2;  
6 TITLE "Cleaned Pre/Post Dataset: OneTwo2";  
7 RUN;
```

Cleaned Pre/Post Dataset: OneTwo2

Obs	ID	Name	Prescore	Postscore
1	11	Joe	19	39
2	12	Ted	21	22
3	13	Darcy	22	42
4	14	Jenny	23	21
5	15	Bill	18	11
6	16	Chris	17	37
7	17	Bill	22	42
8	19	Jeff	24	18

Figure 13: OneTwo2

It is observed that the dataset contains no missing values; therefore, the data are ready to proceed with pre/post-test analysis.