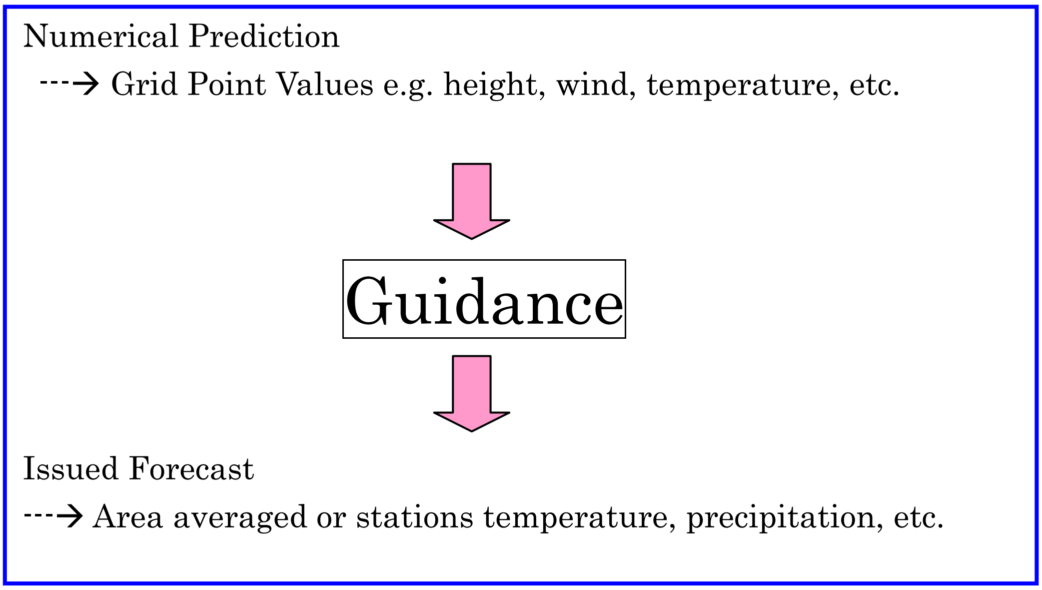
Exercise for Guidance

TCC Training Seminar on Application of Seasonal Forecast GPV Data to Seasonal Forecast Products
18-21 January 2011

Climate Prediction Division, JMA

1. Introduction

Guidance is a statistical downscaling technique from GPV data predicted by numerical model (Fig. 1). Although it is possible to use GPV data for surface elements such as surface temperature and precipitation, there is a possibility to increase accuracy after using guidance. In general, guidance uses some elements such as 500hPa height and 850hPa temperature over the target areas. However, the indices associated with El Niño phenomenon may be more effective in tropics. The purpose of this training is to understand how to make guidance for your countries.



2. Single Regression model

Our situation is that we have a time series of meteorological variable to forecast and a set of time series of other variables obviously related to the former. The former and the latter elements are predictand and predictor, respectively. Our purpose is to predict the future value of predictand using the relationship between predictand and predictors and the present values of predictors.

In order to help the recognition of regression method, let consider the simplest case, that is, single regression. It is a predictive approach using one predictor.

Single regression model is written as

$$Y = ax + b + \varepsilon$$

Y is objective variable (i.e. predictand), **x** is predictor, **a** is regression coefficient, **b** is constant. ε is error term.

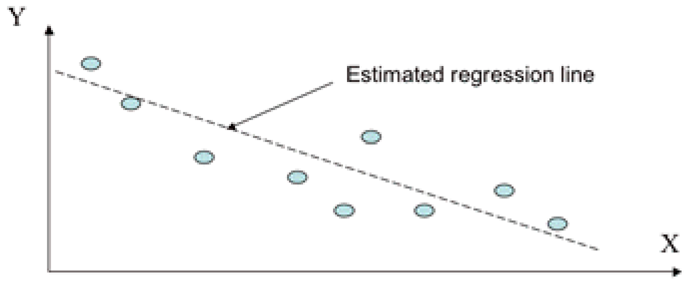


Figure 2 What is Single Regression?

Y is predictand.

X is some other variable that can be used to predict demand.

2-1 Normalization of precipitation

The histogram of temperature is generally normal (Gaussian) distribution but that of precipitation is usually Gamma distribution and has gaps from Gaussian distribution. The error distribution of regression model is assumed normal distribution, so precipitation data is needed to take normalization (Fig. 3). The simplest method is power technique. JMA seasonal forecast guidance is used the power of 1/4 for precipitation and snowfall.

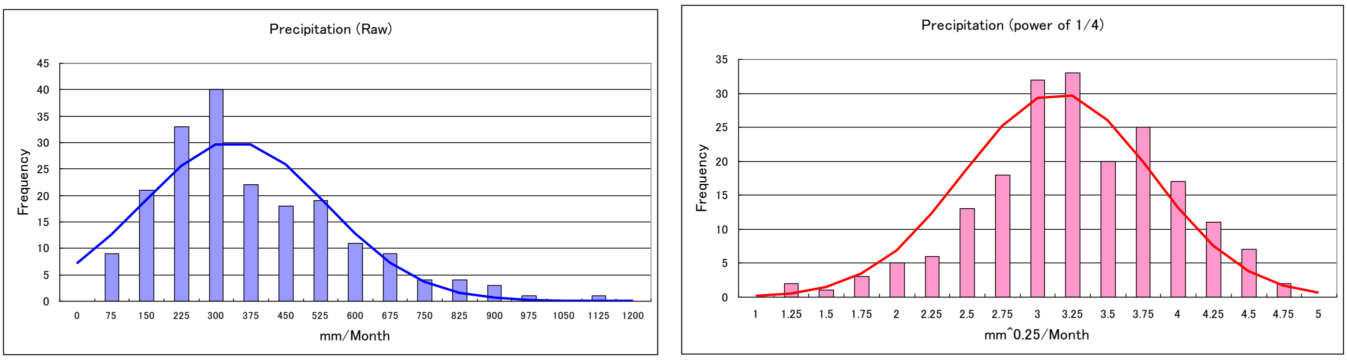


Figure 3 Histogram of precipitation

Raw precipitation data have gap from Gaussian distribution (left).

After taking power of 1/4, the histogram fit Gaussian distribution (right).

Both bold lines indicate Gaussian distribution.

2-2 Methods

Preparation

Observation data

ExerciseForGuidance.xls

GPVdata.xls

Indices.xls

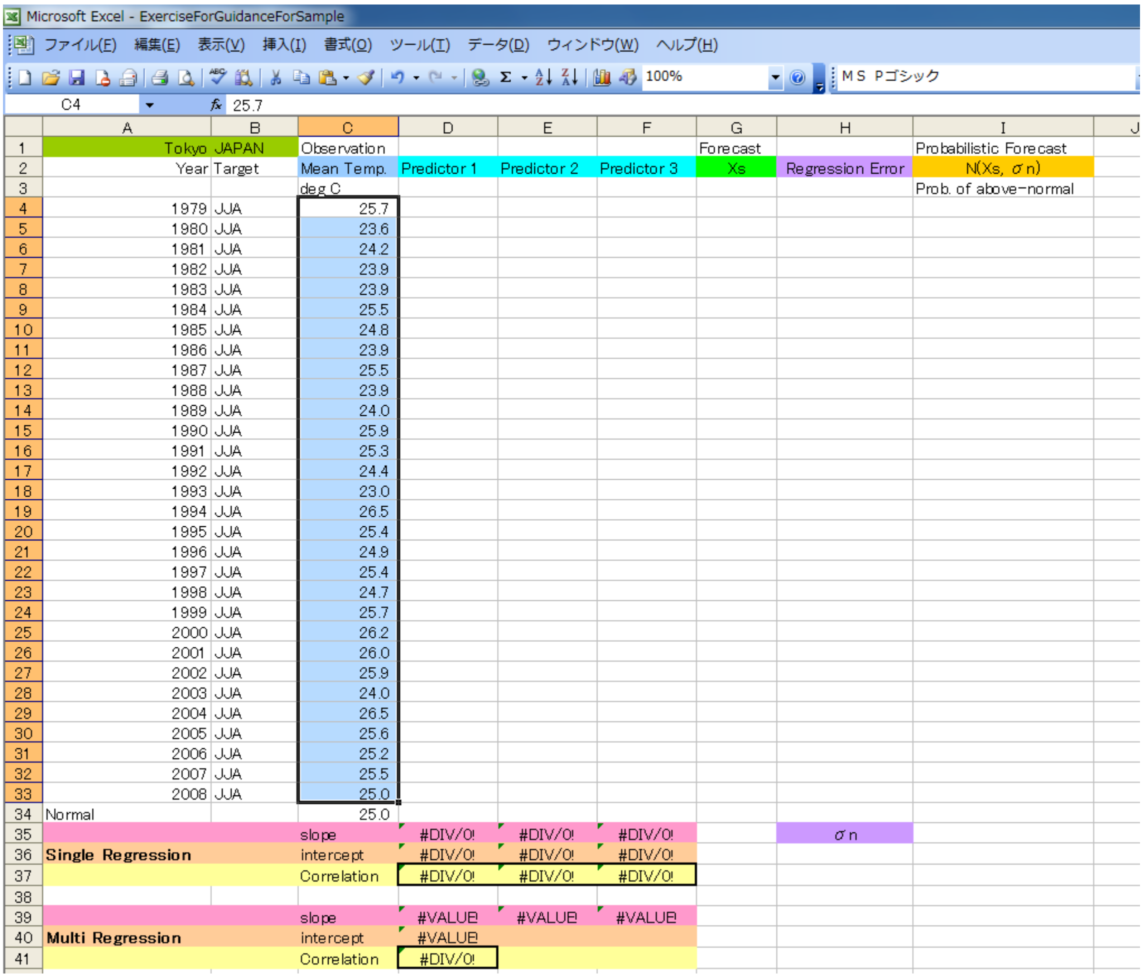
1st step

Open the ExerciseForGuidance.xls.

Paste observation data on a Temperature/Precipitation worksheet.

And input "=AVERAGE(C4:C33)" at C34 to calculate normal.

In case of precipitation, input ="C4^0.25" at D4 to calculate the power of 1/4 and copy D4 and paste D5:D33.



	A	B	C	D	E	F	G	H	I	J
1	Tokyo	JAPAN	Observation				Forecast		Probabilistic Forecast	
2	Year	Target	Mean Temp.	Predictor 1	Predictor 2	Predictor 3	Xs	Regression Error	N(Xs, σn)	
3			deg C						Prob. of above-normal	
4	1979	JJA	25.7							
5	1980	JJA	23.6							
6	1981	JJA	24.2							
7	1982	JJA	23.9							
8	1983	JJA	23.9							
9	1984	JJA	25.5							
10	1985	JJA	24.8							
11	1986	JJA	23.9							
12	1987	JJA	25.5							
13	1988	JJA	23.9							
14	1989	JJA	24.0							
15	1990	JJA	25.9							
16	1991	JJA	25.3							
17	1992	JJA	24.4							
18	1993	JJA	23.0							
19	1994	JJA	26.5							
20	1995	JJA	25.4							
21	1996	JJA	24.9							
22	1997	JJA	25.4							
23	1998	JJA	24.7							
24	1999	JJA	25.7							
25	2000	JJA	26.2							
26	2001	JJA	26.0							
27	2002	JJA	25.9							
28	2003	JJA	24.0							
29	2004	JJA	26.5							
30	2005	JJA	25.6							
31	2006	JJA	25.2							
32	2007	JJA	25.5							
33	2008	JJA	25.0							
34	Normal		25.0							
35			slope	#DIV/0!	#DIV/0!	#DIV/0!		σn		
36	Single Regression		intercept	#DIV/0!	#DIV/0!	#DIV/0!				
37			Correlation	#DIV/0!	#DIV/0!	#DIV/0!				
38										
39			slope	#VALUE!	#VALUE!	#VALUE!				
40	Multi Regression		intercept	#VALUE!						
41			Correlation	#DIV/0!						

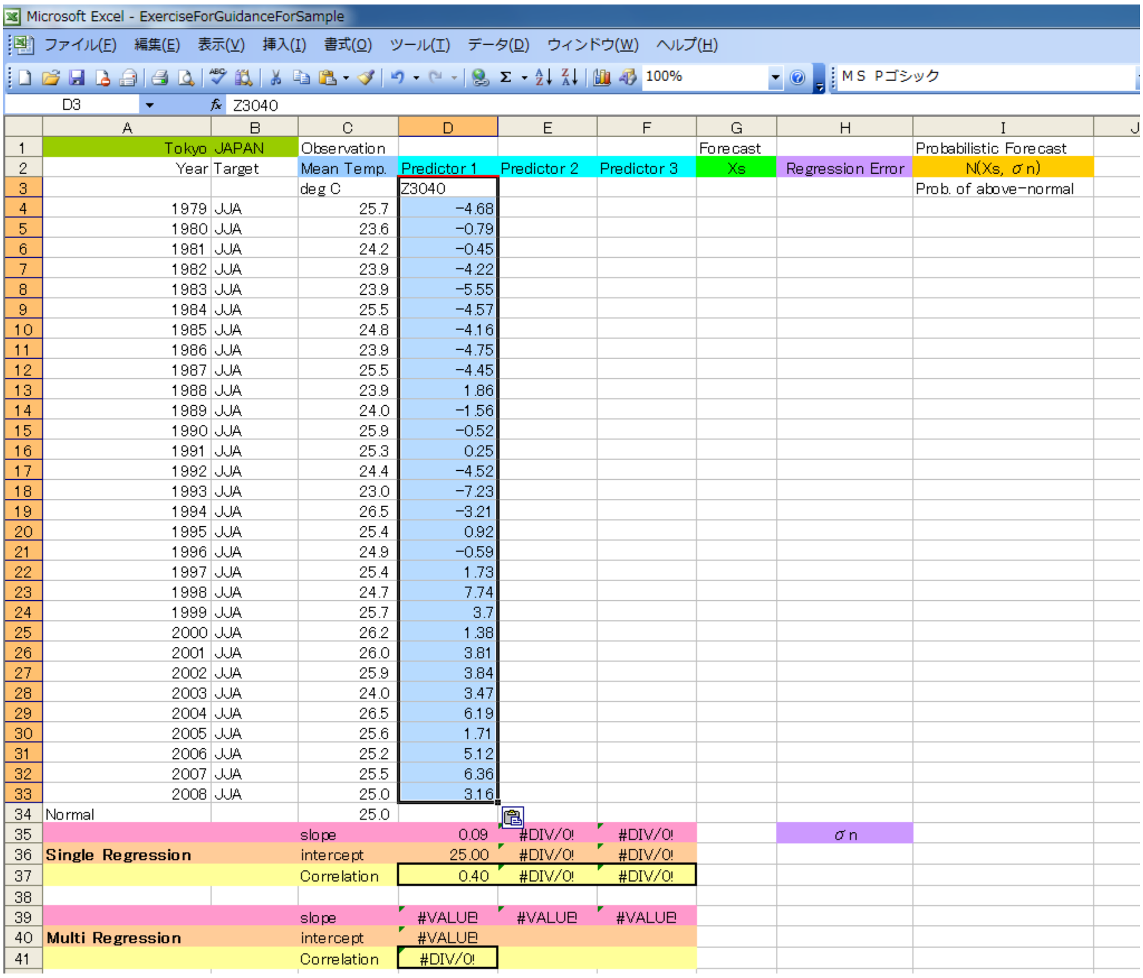
For Example: Paste temperature data on JJA in Tokyo from 1979 to 2008.

2nd step

Open GPVdata.xls and Indices.xls.

Select a predictor and Paste it on D (in case of precipitation, E) line. Try each of predictors until you can find the most effective predictor.

Note: Prepared predictors are anomalies from normal.



	A	B	C	D	E	F	G	H	I
1	Tokyo	JAPAN	Observation				Forecast		Probabilistic Forecast
2	Year	Target	Mean Temp.	Predictor 1	Predictor 2	Predictor 3	Xs	Regression Error	N(Xs, σn)
3			deg C	Z3040					Prob. of above-normal
4	1979	JJA	25.7	-4.68					
5	1980	JJA	23.6	-0.79					
6	1981	JJA	24.2	-0.45					
7	1982	JJA	23.9	-4.22					
8	1983	JJA	23.9	-5.55					
9	1984	JJA	25.5	-4.57					
10	1985	JJA	24.8	-4.16					
11	1986	JJA	23.9	-4.75					
12	1987	JJA	25.5	-4.45					
13	1988	JJA	23.9	1.86					
14	1989	JJA	24.0	-1.56					
15	1990	JJA	25.9	-0.52					
16	1991	JJA	25.3	0.25					
17	1992	JJA	24.4	-4.52					
18	1993	JJA	23.0	-7.23					
19	1994	JJA	26.5	-3.21					
20	1995	JJA	25.4	0.92					
21	1996	JJA	24.9	-0.59					
22	1997	JJA	25.4	1.73					
23	1998	JJA	24.7	7.74					
24	1999	JJA	25.7	3.7					
25	2000	JJA	26.2	1.38					
26	2001	JJA	26.0	3.81					
27	2002	JJA	25.9	3.84					
28	2003	JJA	24.0	3.47					
29	2004	JJA	26.5	6.19					
30	2005	JJA	25.6	1.71					
31	2006	JJA	25.2	5.12					
32	2007	JJA	25.5	6.36					
33	2008	JJA	25.0	3.16					
34	Normal		25.0						
35			slope	0.09	#DIV/0!	#DIV/0!		σn	
36	Single Regression		intercept	25.00	#DIV/0!	#DIV/0!			
37			Correlation	0.40	#DIV/0!	#DIV/0!			
38									
39			slope	#VALUE!	#VALUE!	#VALUE!			
40	Multi Regression		intercept	#VALUE!					
41			Correlation	#DIV/0!					

For Example: Select Zonal Mean height.

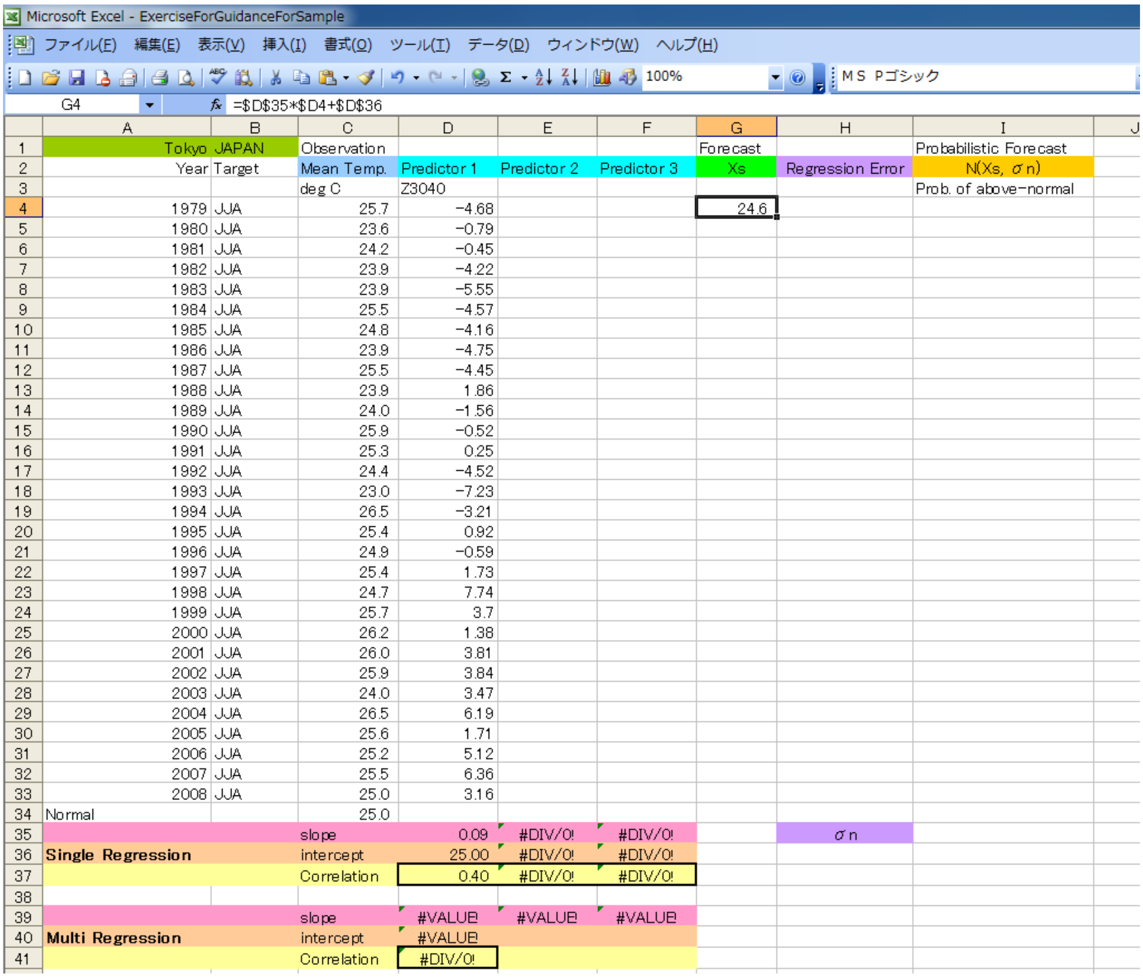
3rd step

Confirm a regression coefficient at D35 and a constant at D36.

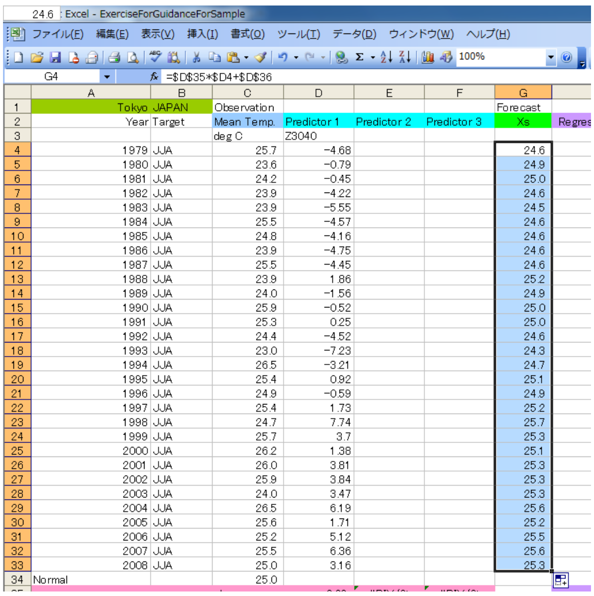
Input a regression equation “=$D$35＊$D4+$D$36” at G4.

Copy G4 and Paste G5:G33.

In case of precipitation, should be E, H instead of D, G.



	A	B	C	D	E	F	G	H	I
1	Tokyo	JAPAN	Observation				Forecast		Probabilistic Forecast
2	Year	Target	Mean Temp.	Predictor 1	Predictor 2	Predictor 3	Xs	Regression Error	N(Xs, σn)
3			deg C	Z3040					Prob. of above-normal
4	1979	JJA	25.7	-4.68			24.6		
5	1980	JJA	23.6	-0.79					
6	1981	JJA	24.2	-0.45					
7	1982	JJA	23.9	-4.22					
8	1983	JJA	23.9	-5.55					
9	1984	JJA	25.5	-4.57					
10	1985	JJA	24.8	-4.16					
11	1986	JJA	23.9	-4.75					
12	1987	JJA	25.5	-4.45					
13	1988	JJA	23.9	1.86					
14	1989	JJA	24.0	-1.56					
15	1990	JJA	25.9	-0.52					
16	1991	JJA	25.3	0.25					
17	1992	JJA	24.4	-4.52					
18	1993	JJA	23.0	-7.23					
19	1994	JJA	26.5	-3.21					
20	1995	JJA	25.4	0.92					
21	1996	JJA	24.9	-0.59					
22	1997	JJA	25.4	1.73					
23	1998	JJA	24.7	7.74					
24	1999	JJA	25.7	3.7					
25	2000	JJA	26.2	1.38					
26	2001	JJA	26.0	3.81					
27	2002	JJA	25.9	3.84					
28	2003	JJA	24.0	3.47					
29	2004	JJA	26.5	6.19					
30	2005	JJA	25.6	1.71					
31	2006	JJA	25.2	5.12					
32	2007	JJA	25.5	6.36					
33	2008	JJA	25.0	3.16					
34	Normal		25.0						
35			slope	0.09	#DIV/0!	#DIV/0!		σn	
36	Single Regression		intercept	25.00	#DIV/0!	#DIV/0!			
37			Correlation	0.40	#DIV/0!	#DIV/0!			
38									
39			slope	#VALUE!	#VALUE!	#VALUE!			
40	Multi Regression		intercept	#VALUE!					
41			Correlation	#DIV/0!					



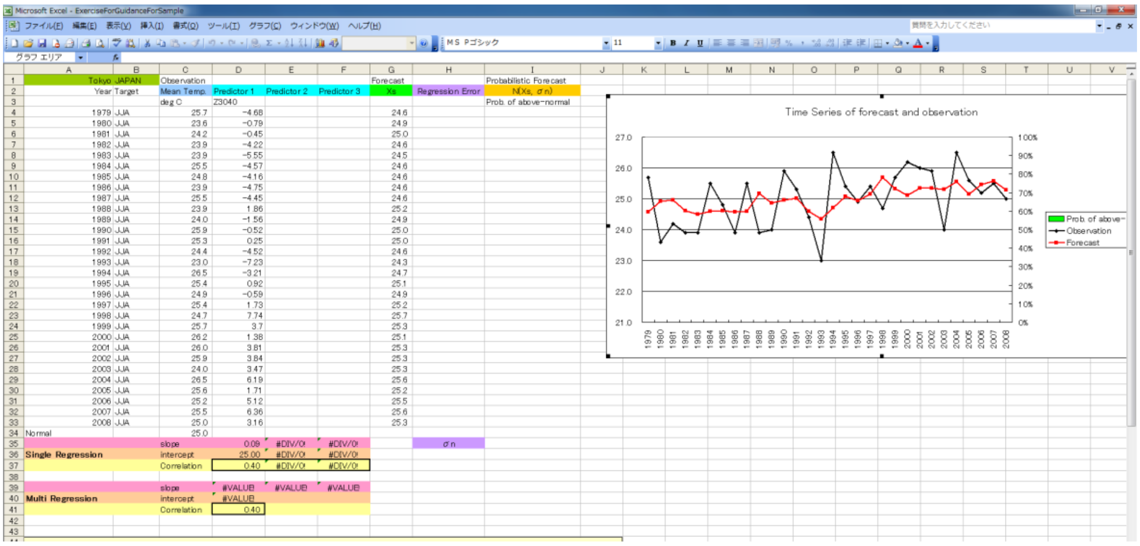
	A	B	C	D	E	F	G
1	Tokyo	JAPAN	Observation				Forecast
2	Year	Target	Mean Temp.	Predictor 1	Predictor 2	Predictor 3	Xs
3			deg C	Z3040			
4	1979	JJA	25.7	-4.68			24.6
5	1980	JJA	23.6	-0.79			24.9
6	1981	JJA	24.2	-0.45			25.0
7	1982	JJA	23.9	-4.22			24.6
8	1983	JJA	23.9	-5.55			24.5
9	1984	JJA	25.5	-4.57			24.6
10	1985	JJA	24.8	-4.16			24.6
11	1986	JJA	23.9	-4.75			24.6
12	1987	JJA	25.5	-4.45			24.6
13	1988	JJA	23.9	1.86			25.2
14	1989	JJA	24.0	-1.56			24.9
15	1990	JJA	25.9	-0.52			25.0
16	1991	JJA	25.3	0.25			25.0
17	1992	JJA	24.4	-4.52			24.6
18	1993	JJA	23.0	-7.23			24.3
19	1994	JJA	26.5	-3.21			24.7
20	1995	JJA	25.4	0.92			25.1
21	1996	JJA	24.9	-0.59			24.9
22	1997	JJA	25.4	1.73			25.2
23	1998	JJA	24.7	7.74			25.7
24	1999	JJA	25.7	3.7			25.3
25	2000	JJA	26.2	1.38			25.1
26	2001	JJA	26.0	3.81			25.3
27	2002	JJA	25.9	3.84			25.3
28	2003	JJA	24.0	3.47			25.3
29	2004	JJA	26.5	6.19			25.6
30	2005	JJA	25.6	1.71			25.2
31	2006	JJA	25.2	5.12			25.5
32	2007	JJA	25.5	6.36			25.6
33	2008	JJA	25.0	3.16			25.3
34	Normal		25.0				

4th step

This is the end of single regression model.

You can see a time series line chart of forecast and observation.

Confirm an anomaly correlation coefficient at D37.



	A	B	C	D	E	F	G	H	I
1	Tokyo JAPAN		Observation				Forecast		Probabilistic Forecast
2	Year	Target	Mean Temp.	Predictor 1	Predictor 2	Predictor 3	Xs	Regression Error	N(Xs, σn)
3			deg C	Z3040					Prob. of above-normal
4	1979	JJA	25.7	-4.68			24.6		
5	1980	JJA	23.6	-0.79			24.9		
6	1981	JJA	24.2	-0.45			25.0		
7	1982	JJA	23.9	-4.22			24.6		
8	1983	JJA	23.9	-5.55			24.5		
9	1984	JJA	25.5	-4.57			24.6		
10	1985	JJA	24.8	-4.16			24.6		
11	1986	JJA	23.9	-4.75			24.6		
12	1987	JJA	25.5	-4.45			24.6		
13	1988	JJA	23.9	1.86			25.2		
14	1989	JJA	24.0	-1.56			24.9		
15	1990	JJA	25.9	-0.52			25.0		
16	1991	JJA	25.3	0.25			25.0		
17	1992	JJA	24.4	-4.52			24.6		
18	1993	JJA	23.0	-7.23			24.3		
19	1994	JJA	26.5	-3.21			24.7		
20	1995	JJA	25.4	0.92			25.1		
21	1996	JJA	24.9	-0.59			24.9		
22	1997	JJA	25.4	1.73			25.2		
23	1998	JJA	24.7	7.74			25.7		
24	1999	JJA	25.7	3.7			25.3		
25	2000	JJA	26.2	1.38			25.1		
26	2001	JJA	26.0	3.81			25.3		
27	2002	JJA	25.9	3.84			25.3		
28	2003	JJA	24.0	3.47			25.3		
29	2004	JJA	26.5	6.19			25.6		
30	2005	JJA	25.6	1.71			25.2		
31	2006	JJA	25.2	5.12			25.5		
32	2007	JJA	25.5	6.36			25.6		
33	2008	JJA	25.0	3.16			25.3		
34	Normal		25.0						
35			slope	0.09	#DIV/0!	#DIV/0!		σn	
36	Single Regression		intercept	25.00	#DIV/0!	#DIV/0!			
37			Correlation	0.40	#DIV/0!	#DIV/0!			
38									
39			slope	#VALUE!	#VALUE!	#VALUE!			
40	Multi Regression		intercept	#VALUE!					
41			Correlation	0.40					

2-3 Questions

What predictor do you select?

Can you get an accuracy guidance?

How does its guidance predict the hottest/coldest/drought/wet year in your country?

3. Multiple Regression model

The multiple regression model assumes predictand is sum of a linear combination of predictors.

Multiple regression model is written as

$$Y = a_t x_t + b + \varepsilon_t \qquad t=1,2,3,\ldots\ldots,n$$

Y is objective variable (i.e. predictand), **x** is predictors, **a** is regression coefficients, **b** is constant. ε is error term.

3-1 Methods

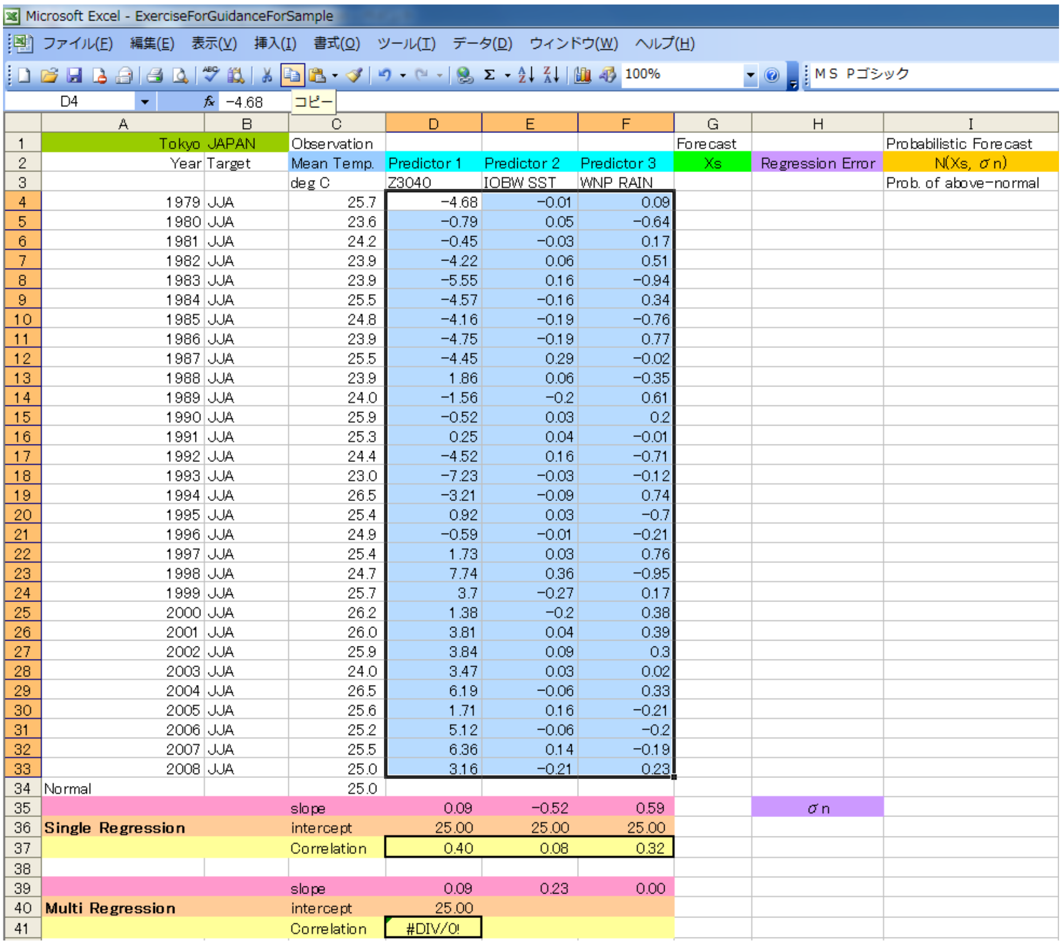
Preparation and 1st step is same as single regression model.

2nd step

Open GPVdata.xls and Indices.xls.

Select some predictors and Paste it on D,E,F line. Try each of predictors until you can find the most effective combination of predictors while confirming anomaly correlation coefficient at D37,E37,F37.

In case of precipitation, should be E, F, G line instead of D, E, F line.



	A	B	C	D	E	F	G	H	I
1	Tokyo	JAPAN	Observation				Forecast		Probabilistic Forecast
2	Year	Target	Mean Temp.	Predictor 1	Predictor 2	Predictor 3	Xs	Regression Error	N(Xs, σn)
3			deg C	Z3040	IOBW SST	WNP RAIN			Prob. of above-normal
4	1979	JJA	25.7	-4.68	-0.01	0.09			
5	1980	JJA	23.6	-0.79	0.05	-0.64			
6	1981	JJA	24.2	-0.45	-0.03	0.17			
7	1982	JJA	23.9	-4.22	0.06	0.51			
8	1983	JJA	23.9	-5.55	0.16	-0.94			
9	1984	JJA	25.5	-4.57	-0.16	0.34			
10	1985	JJA	24.8	-4.16	-0.19	-0.76			
11	1986	JJA	23.9	-4.75	-0.19	0.77			
12	1987	JJA	25.5	-4.45	0.29	-0.02			
13	1988	JJA	23.9	1.86	0.06	-0.35			
14	1989	JJA	24.0	-1.56	-0.2	0.61			
15	1990	JJA	25.9	-0.52	0.03	0.2			
16	1991	JJA	25.3	0.25	0.04	-0.01			
17	1992	JJA	24.4	-4.52	0.16	-0.71			
18	1993	JJA	23.0	-7.23	-0.03	-0.12			
19	1994	JJA	26.5	-3.21	-0.09	0.74			
20	1995	JJA	25.4	0.92	0.03	-0.7			
21	1996	JJA	24.9	-0.59	-0.01	-0.21			
22	1997	JJA	25.4	1.73	0.03	0.76			
23	1998	JJA	24.7	7.74	0.36	-0.95			
24	1999	JJA	25.7	3.7	-0.27	0.17			
25	2000	JJA	26.2	1.38	-0.2	0.38			
26	2001	JJA	26.0	3.81	0.04	0.39			
27	2002	JJA	25.9	3.84	0.09	0.3			
28	2003	JJA	24.0	3.47	0.03	0.02			
29	2004	JJA	26.5	6.19	-0.06	0.33			
30	2005	JJA	25.6	1.71	0.16	-0.21			
31	2006	JJA	25.2	5.12	-0.06	-0.2			
32	2007	JJA	25.5	6.36	0.14	-0.19			
33	2008	JJA	25.0	3.16	-0.21	0.23			
34	Normal		25.0						
35			slope	0.09	-0.52	0.59		σn	
36	**Single Regression**		intercept	25.00	25.00	25.00			
37			Correlation	0.40	0.08	0.32			
38									
39			slope	0.09	0.23	0.00			
40	**Multi Regression**		intercept	25.00					
41			Correlation	#DIV/0!					

For Example: Zonal mean height and Indian Ocean SST and Western North Pacific Precipitation are selected as predictors for temperature in Tokyo.

3rd step

If you select two predictors, input "=$D$39＊$D4+$E$39＊$E4+$D$40" at G4.

If you select three predictors, input "=$D$39＊$D4+$E$39＊$E4+$F$39＊$F4+$D$40" at G4.

Copy G4 and Paste G5:G33.

And you can confirm an anomaly correlation coefficient at D41.

If the value of E41 is less than those of D37, E37, F37, let's try the other combination of predictors until you can get high accuracy multiple regression equation.

In case of precipitation, should be E, F, G, H line instead of D, E, F, G line.

Microsoft Excel - ExerciseForGuidanceForSample

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

G5 | =$D$39*$D5+$E$39*$E5+$F$39*$F5+$D$40

	A	B	C	D	E	F	G	H	I
1	Tokyo	JAPAN	Observation				Forecast		Probabilistic Foreca
2	Year	Target	Mean Temp.	Predictor 1	Predictor 2	Predictor 3	Xs	Regression Error	N(Xs, σn)
3			deg C	Z3040	NINOWEST	WNP RAIN			Prob. of above-nor
4	1979	JJA	25.7	-4.68	-0.03	0.09	24.7		
5	1980	JJA	23.6	-0.79	-0.1	-0.64	24.8		
6	1981	JJA	24.2	-0.45	0.04	0.17	25.0		
7	1982	JJA	23.9	-4.22	-0.17	0.51	24.5		
8	1983	JJA	23.9	-5.55	-0.14	-0.94	24.5		
9	1984	JJA	25.5	-4.57	0.05	0.34	24.8		
10	1985	JJA	24.8	-4.16	0	-0.76	24.7		
11	1986	JJA	23.9	-4.75	-0.02	0.77	24.7		
12	1987	JJA	25.5	-4.45	-0.13	-0.02	24.6		
13	1988	JJA	23.9	1.86	0.07	-0.35	25.2		
14	1989	JJA	24.0	-1.56	0.01	0.61	24.9		
15	1990	JJA	25.9	-0.52	0.01	0.2	25.0		
16	1991	JJA	25.3	0.25	-0.17	-0.01	24.8		
17	1992	JJA	24.4	-4.52	-0.23	-0.71	24.5		
18	1993	JJA	23.0	-7.23	-0.47	-0.12	24.0		
19	1994	JJA	26.5	-3.21	-0.05	0.74	24.7		
20	1995	JJA	25.4	0.92	0	-0.7	25.1		
21	1996	JJA	24.9	-0.59	0.17	-0.21	25.2		
22	1997	JJA	25.4	1.73	-0.31	0.76	24.7		
23	1998	JJA	24.7	7.74	0.12	-0.95	25.6		
24	1999	JJA	25.7	3.7	0.03	0.17	25.3		
25	2000	JJA	26.2	1.38	0.15	0.38	25.3		
26	2001	JJA	26.0	3.81	0.27	0.39	25.5		
27	2002	JJA	25.9	3.84	0.01	0.3	25.2		
28	2003	JJA	24.0	3.47	0.26	0.02	25.5		
29	2004	JJA	26.5	6.19	0.08	0.33	25.5		
30	2005	JJA	25.6	1.71	0.19	-0.21	25.3		
31	2006	JJA	25.2	5.12	0.15	-0.2	25.5		
32	2007	JJA	25.5	6.36	0.19	-0.19	25.6		
33	2008	JJA	25.0	3.16	0.03	0.23	25.2		
34	Normal		25.0						
35			slope	0.09	2.12	0.59		σn	
36	Single Regression		intercept	25.00	25.00	25.00			
37			Correlation	0.40	0.38	0.32			
38									
39			slope	0.06	1.15	0.00			
40	Multiple Regression		intercept	25.00					
41			Correlation	0.43					

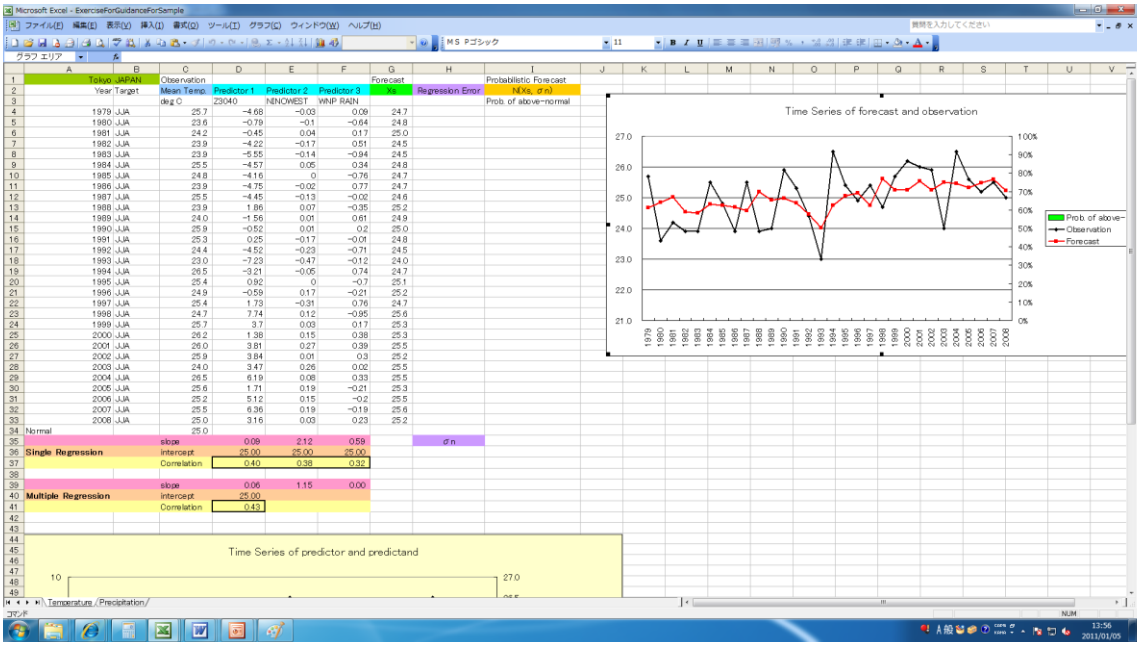
For Example: Selecting NINOWEST SST instead of Indian Ocean SST make the correlation coefficient increase.

4th step

This is the end of multiple regression model.

You can see a time series line chart of forecast and observation.

In case of precipitation, input "=H4^4" at I4 to power of 4, and copy I4 and paste I5:I33.



3-2 Questions

What predictors do you select?

Can you get more accuracy guidance than single regression model?

How does its guidance predict the hottest/coldest/drought/wet year in your country?

3-3 Reference

See the first reference materials about regression model in detail.

4. Probabilistic Forecast

Seasonal forecast has uncertainty due to chaotic character of the atmospheric flow. Therefore it is necessary to take into account uncertainty of forecast. To do this, the best method is the probabilistic forecast. Here we present probabilistic forecast issued by the Japan Meteorological Agency (JMA) and verification of probabilistic forecast.

The Probability Density Function (PDF) is assumed as normal distribution (Gaussian distribution) with mean x_s and standard deviation σ_n (Fig. 4).
The mean x_s is predicted by single/multiple regression model and standard deviation σ_n is assumed as root mean square error of regression model.

JMA has issued three-categorical probabilistic forecast based on this guidance around 25th every month (Fig. 5).

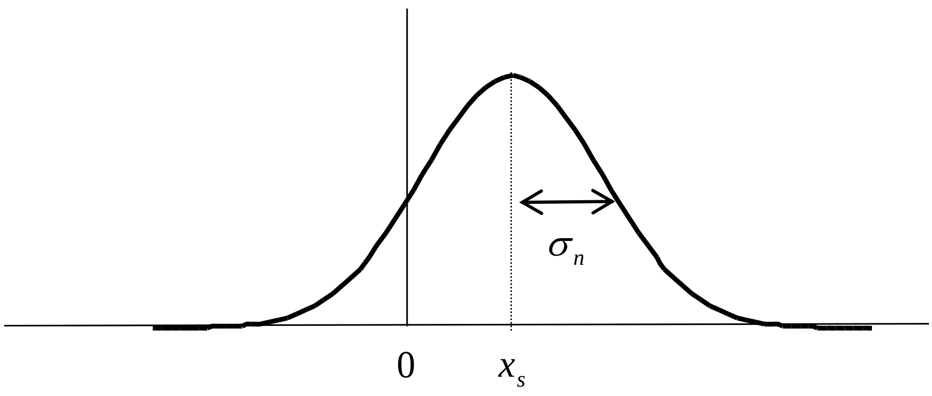
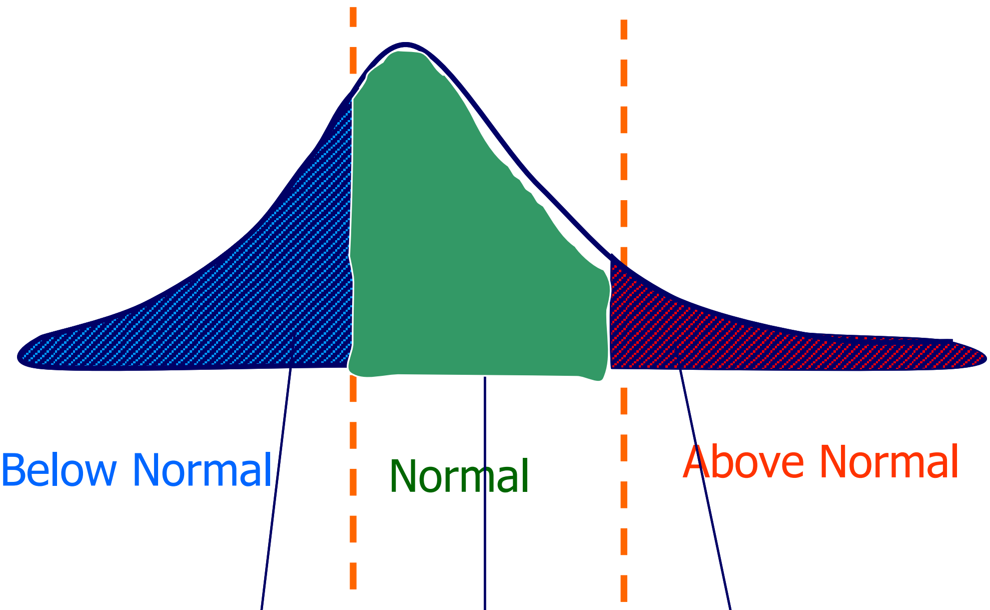


Figure 4 Schematic diagram of forecast probability density function (PDF)



Below Nomal	Normal	Above Normal
40%	40%	20%

Figure 5 Schematic diagram of three categorical forecast
Dash lines indicate threshold value.

4-1 Methods

Preparation

ExerciseForGuidance.xls

Note the line for precipitation same as multiple regression model.

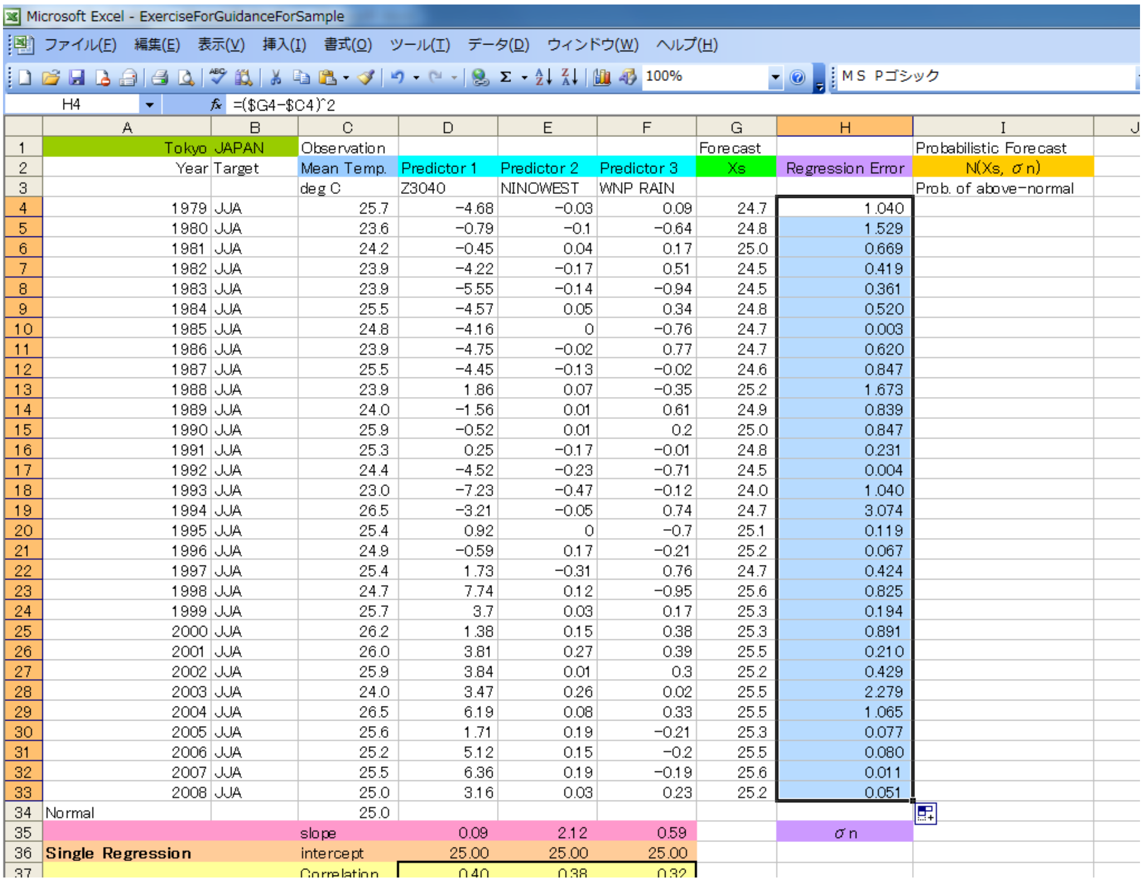
1st step

To calculate square of regression error, input "=($G4-$C4)^2" at H4.

$\varepsilon^2 = \{y - (ax + b)\}^2$

Copy H4 and Pate H5:H33.

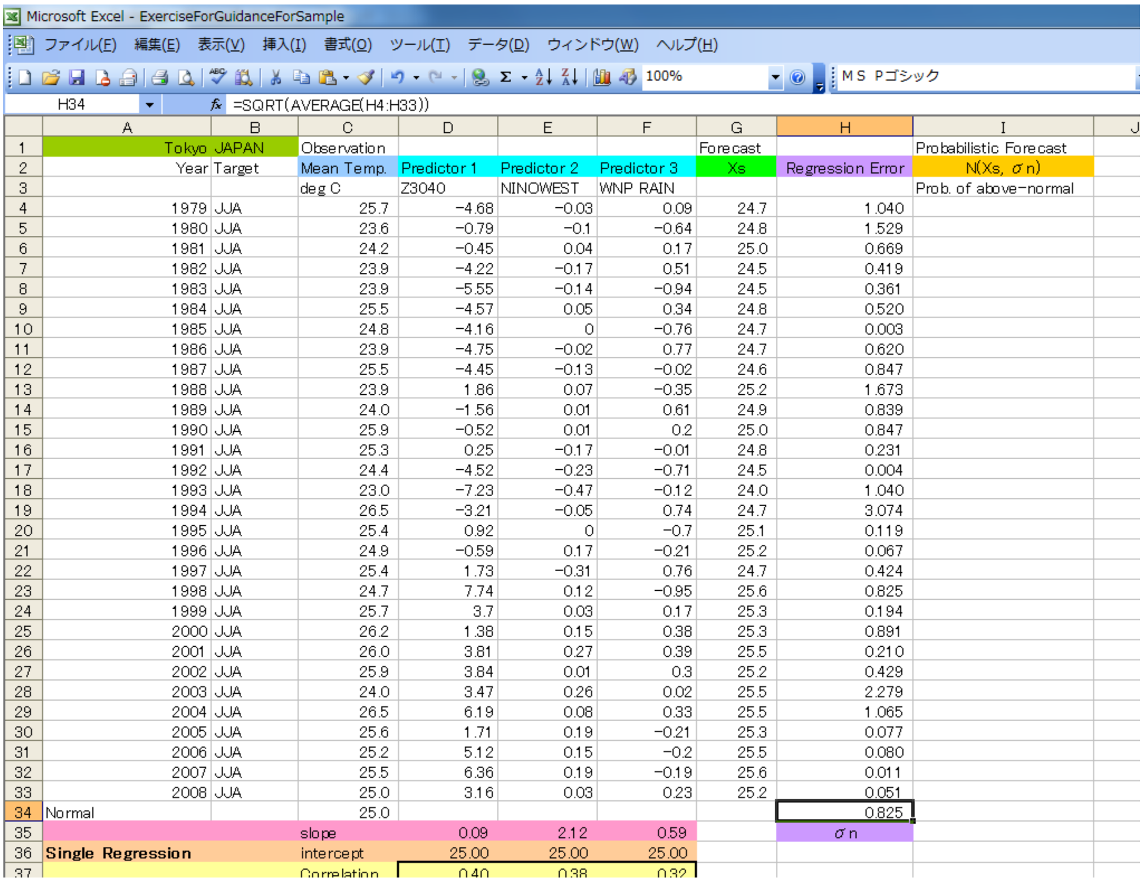
In case of precipitation, should be I, J line instead of G, H line.



	A	B	C	D	E	F	G	H	I
1	Tokyo	JAPAN	Observation				Forecast		Probabilistic Forecast
2	Year	Target	Mean Temp.	Predictor 1	Predictor 2	Predictor 3	Xs	Regression Error	N(Xs, σn)
3			deg C	Z3040	NINOWEST	WNP RAIN			Prob. of above-normal
4	1979	JJA	25.7	−4.68	−0.03	0.09	24.7	1.040	
5	1980	JJA	23.6	−0.79	−0.1	−0.64	24.8	1.529	
6	1981	JJA	24.2	−0.45	0.04	0.17	25.0	0.669	
7	1982	JJA	23.9	−4.22	−0.17	0.51	24.5	0.419	
8	1983	JJA	23.9	−5.55	−0.14	−0.94	24.5	0.361	
9	1984	JJA	25.5	−4.57	0.05	0.34	24.8	0.520	
10	1985	JJA	24.8	−4.16	0	−0.76	24.7	0.003	
11	1986	JJA	23.9	−4.75	−0.02	0.77	24.7	0.620	
12	1987	JJA	25.5	−4.45	−0.13	−0.02	24.6	0.847	
13	1988	JJA	23.9	1.86	0.07	−0.35	25.2	1.673	
14	1989	JJA	24.0	−1.56	0.01	0.61	24.9	0.839	
15	1990	JJA	25.9	−0.52	0.01	0.2	25.0	0.847	
16	1991	JJA	25.3	0.25	−0.17	−0.01	24.8	0.231	
17	1992	JJA	24.4	−4.52	−0.23	−0.71	24.5	0.004	
18	1993	JJA	23.0	−7.23	−0.47	−0.12	24.0	1.040	
19	1994	JJA	26.5	−3.21	−0.05	0.74	24.7	3.074	
20	1995	JJA	25.4	0.92	0	−0.7	25.1	0.119	
21	1996	JJA	24.9	−0.59	0.17	−0.21	25.2	0.067	
22	1997	JJA	25.4	1.73	−0.31	0.76	24.7	0.424	
23	1998	JJA	24.7	7.74	0.12	−0.95	25.6	0.825	
24	1999	JJA	25.7	3.7	0.03	0.17	25.3	0.194	
25	2000	JJA	26.2	1.38	0.15	0.38	25.3	0.891	
26	2001	JJA	26.0	3.81	0.27	0.39	25.5	0.210	
27	2002	JJA	25.9	3.84	0.01	0.3	25.2	0.429	
28	2003	JJA	24.0	3.47	0.26	0.02	25.5	2.279	
29	2004	JJA	26.5	6.19	0.08	0.33	25.5	1.065	
30	2005	JJA	25.6	1.71	0.19	−0.21	25.3	0.077	
31	2006	JJA	25.2	5.12	0.15	−0.2	25.5	0.080	
32	2007	JJA	25.5	6.36	0.19	−0.19	25.6	0.011	
33	2008	JJA	25.0	3.16	0.03	0.23	25.2	0.051	
34	Normal		25.0						
35			slope	0.09	2.12	0.59		σn	
36	**Single Regression**		intercept	25.00	25.00	25.00			
37			Correlation	0.40	0.38	0.32			

2nd step

To calculate Root Mean square error, input “=SQRT(AVERAGE(H4:H33))” at H34. This value is used as standard deviation of normal distribution.



	A	B	C	D	E	F	G	H	I
1	Tokyo	JAPAN	Observation				Forecast		Probabilistic Forecast
2	Year	Target	Mean Temp.	Predictor 1	Predictor 2	Predictor 3	Xs	Regression Error	N(Xs, σn)
3			deg C	Z3040	NINOWEST	WNP RAIN			Prob. of above-normal
4	1979	JJA	25.7	-4.68	-0.03	0.09	24.7	1.040	
5	1980	JJA	23.6	-0.79	-0.1	-0.64	24.8	1.529	
6	1981	JJA	24.2	-0.45	0.04	0.17	25.0	0.669	
7	1982	JJA	23.9	-4.22	-0.17	0.51	24.5	0.419	
8	1983	JJA	23.9	-5.55	-0.14	-0.94	24.5	0.361	
9	1984	JJA	25.5	-4.57	0.05	0.34	24.8	0.520	
10	1985	JJA	24.8	-4.16	0	-0.76	24.7	0.003	
11	1986	JJA	23.9	-4.75	-0.02	0.77	24.7	0.620	
12	1987	JJA	25.5	-4.45	-0.13	-0.02	24.6	0.847	
13	1988	JJA	23.9	1.86	0.07	-0.35	25.2	1.673	
14	1989	JJA	24.0	-1.56	0.01	0.61	24.9	0.839	
15	1990	JJA	25.9	-0.52	0.01	0.2	25.0	0.847	
16	1991	JJA	25.3	0.25	-0.17	-0.01	24.8	0.231	
17	1992	JJA	24.4	-4.52	-0.23	-0.71	24.5	0.004	
18	1993	JJA	23.0	-7.23	-0.47	-0.12	24.0	1.040	
19	1994	JJA	26.5	-3.21	-0.05	0.74	24.7	3.074	
20	1995	JJA	25.4	0.92	0	-0.7	25.1	0.119	
21	1996	JJA	24.9	-0.59	0.17	-0.21	25.2	0.067	
22	1997	JJA	25.4	1.73	-0.31	0.76	24.7	0.424	
23	1998	JJA	24.7	7.74	0.12	-0.95	25.6	0.825	
24	1999	JJA	25.7	3.7	0.03	0.17	25.3	0.194	
25	2000	JJA	26.2	1.38	0.15	0.38	25.3	0.891	
26	2001	JJA	26.0	3.81	0.27	0.39	25.5	0.210	
27	2002	JJA	25.9	3.84	0.01	0.3	25.2	0.429	
28	2003	JJA	24.0	3.47	0.26	0.02	25.5	2.279	
29	2004	JJA	26.5	6.19	0.08	0.33	25.5	1.065	
30	2005	JJA	25.6	1.71	0.19	-0.21	25.3	0.077	
31	2006	JJA	25.2	5.12	0.15	-0.2	25.5	0.080	
32	2007	JJA	25.5	6.36	0.19	-0.19	25.6	0.011	
33	2008	JJA	25.0	3.16	0.03	0.23	25.2	0.051	
34	Normal		25.0					0.825	
35			slope	0.09	2.12	0.59		σn	
36	Single Regression		intercept	25.00	25.00	25.00			
37			Correlation	0.40	0.38	0.32			

3rd step

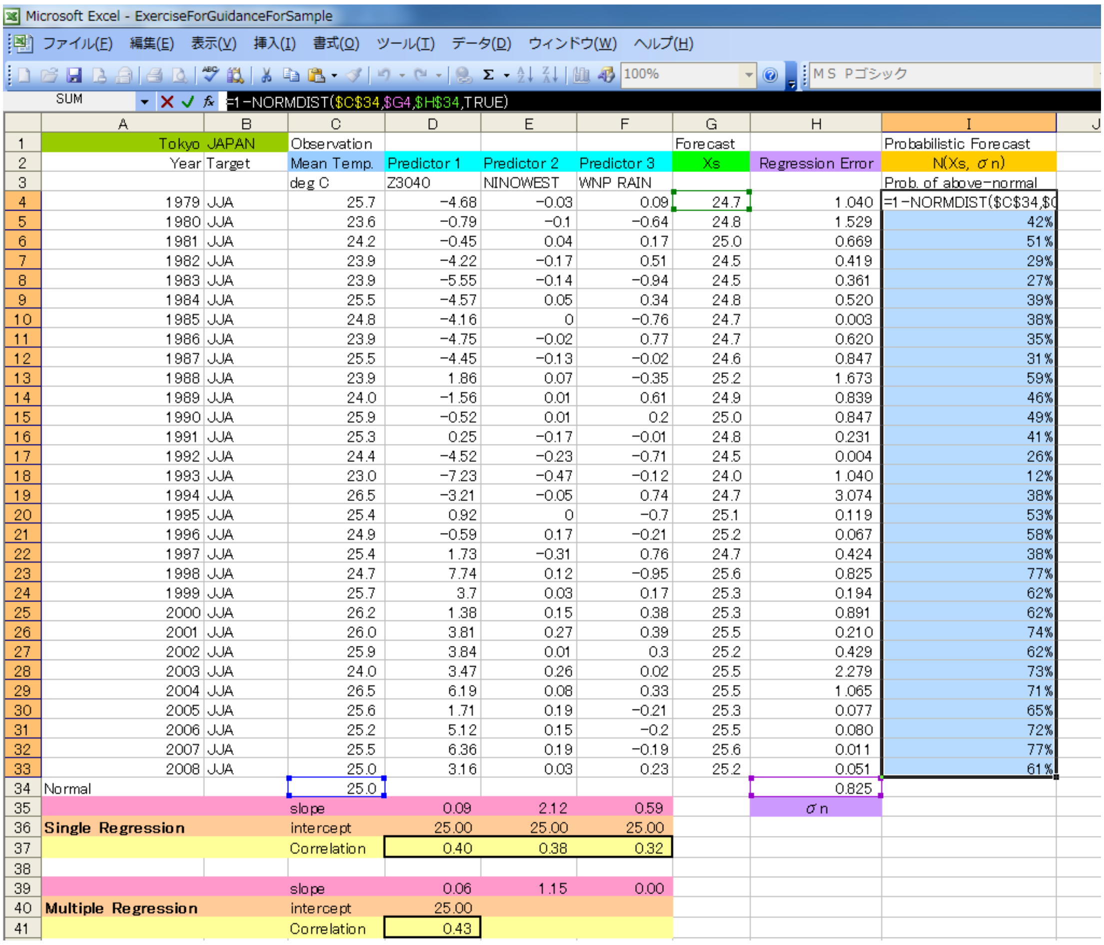
Assuming normal distribution $N(X_s, \sigma_n)$,

input "=1-NORMDIST($C$34,$G4,$H$34,TRUE)" at I4.

Copy I4 and paste I5:I33.

The values of I line indicate the probability of above-normal.

In case of precipitation, should be K line instead of I line.


Microsoft Excel - ExerciseForGuidanceForSample

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

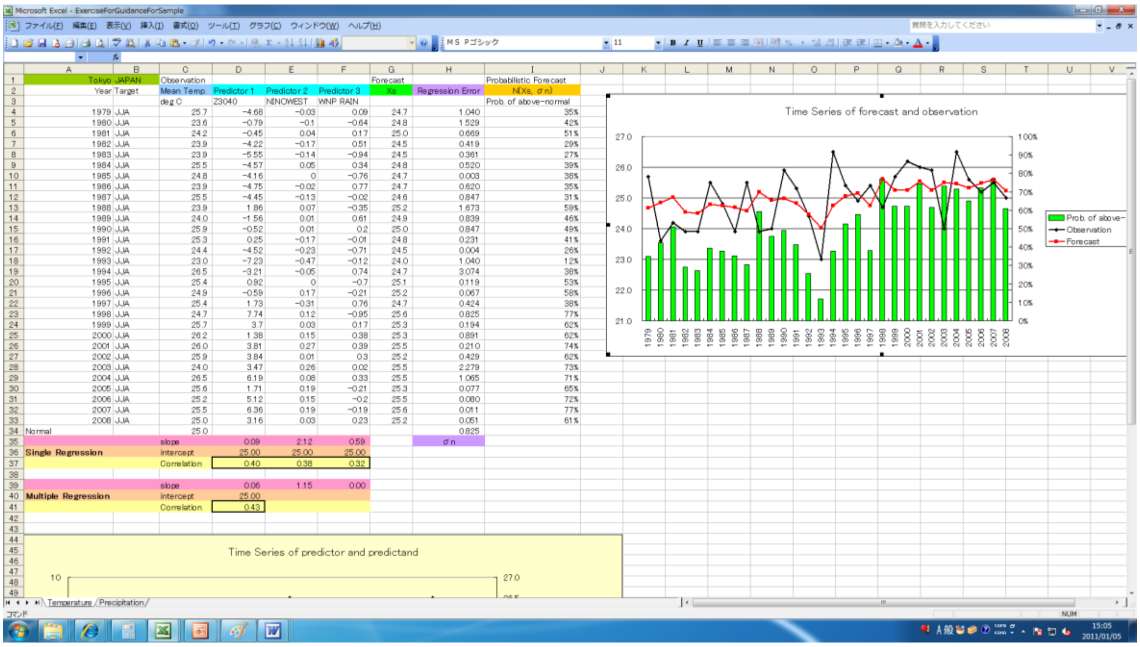
SUM =1-NORMDIST($C$34,$G4,$H$34,TRUE)

	A	B	C	D	E	F	G	H	I
1	Tokyo	JAPAN	Observation				Forecast		Probabilistic Forecast
2	Year	Target	Mean Temp.	Predictor 1	Predictor 2	Predictor 3	Xs	Regression Error	N(Xs, σn)
3			deg C	Z3040	NINOWEST	WNP RAIN			Prob. of above-normal
4	1979	JJA	25.7	-4.68	-0.03	0.09	24.7	1.040	=1-NORMDIST($C$34,$G
5	1980	JJA	23.6	-0.79	-0.1	-0.64	24.8	1.529	42%
6	1981	JJA	24.2	-0.45	0.04	0.17	25.0	0.669	51%
7	1982	JJA	23.9	-4.22	-0.17	0.51	24.5	0.419	29%
8	1983	JJA	23.9	-5.55	-0.14	-0.94	24.5	0.361	27%
9	1984	JJA	25.5	-4.57	0.05	0.34	24.8	0.520	39%
10	1985	JJA	24.8	-4.16	0	-0.76	24.7	0.003	38%
11	1986	JJA	23.9	-4.75	-0.02	0.77	24.7	0.620	35%
12	1987	JJA	25.5	-4.45	-0.13	-0.02	24.6	0.847	31%
13	1988	JJA	23.9	1.86	0.07	-0.35	25.2	1.673	59%
14	1989	JJA	24.0	-1.56	0.01	0.61	24.9	0.839	46%
15	1990	JJA	25.9	-0.52	0.01	0.2	25.0	0.847	49%
16	1991	JJA	25.3	0.25	-0.17	-0.01	24.8	0.231	41%
17	1992	JJA	24.4	-4.52	-0.23	-0.71	24.5	0.004	26%
18	1993	JJA	23.0	-7.23	-0.47	-0.12	24.0	1.040	12%
19	1994	JJA	26.5	-3.21	-0.05	0.74	24.7	3.074	38%
20	1995	JJA	25.4	0.92	0	-0.7	25.1	0.119	53%
21	1996	JJA	24.9	-0.59	0.17	-0.21	25.2	0.067	58%
22	1997	JJA	25.4	1.73	-0.31	0.76	24.7	0.424	38%
23	1998	JJA	24.7	7.74	0.12	-0.95	25.6	0.825	77%
24	1999	JJA	25.7	3.7	0.03	0.17	25.3	0.194	62%
25	2000	JJA	26.2	1.38	0.15	0.38	25.3	0.891	62%
26	2001	JJA	26.0	3.81	0.27	0.39	25.5	0.210	74%
27	2002	JJA	25.9	3.84	0.01	0.3	25.2	0.429	62%
28	2003	JJA	24.0	3.47	0.26	0.02	25.5	2.279	73%
29	2004	JJA	26.5	6.19	0.08	0.33	25.5	1.065	71%
30	2005	JJA	25.6	1.71	0.19	-0.21	25.3	0.077	65%
31	2006	JJA	25.2	5.12	0.15	-0.2	25.5	0.080	72%
32	2007	JJA	25.5	6.36	0.19	-0.19	25.6	0.011	77%
33	2008	JJA	25.0	3.16	0.03	0.23	25.2	0.051	61%
34	Normal		25.0					0.825	
35			slope	0.09	2.12	0.59		σn	
36	Single Regression		intercept	25.00	25.00	25.00			
37			Correlation	0.40	0.38	0.32			
38									
39			slope	0.06	1.15	0.00			
40	Multiple Regression		intercept	25.00					
41			Correlation	0.43					

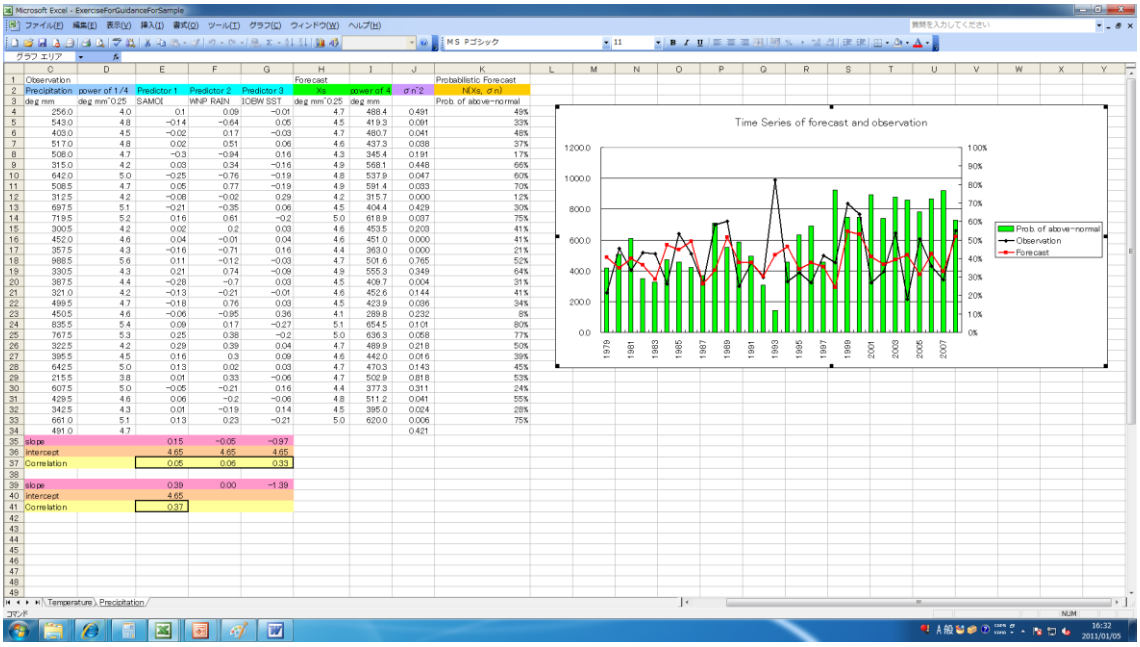
4th step

This is the end of probabilistic forecast.

You can see a time series bar chart of probabilistic forecast.



For Example: Completion of probabilistic forecast for temperature.



For Example: Completion of probabilistic forecast for precipitation.

4-2 Question

What is the difference of probability between temperature and precipitation?

4-3 Reference

See the second reference materials about seasonal forecast and predictability.

The following are additional exercises

5. Evaluation of Probability forecast

In this section, we will present two techniques for evaluating probability forecasting: one is the reliability diagram, which shows occurrence frequencies by forecast probability, and the Brier score used to evaluate the forecasts; and the other is the ROC (relative operating characteristic) curve, which is drawn based on “hit rates” and “false alarm rates” that are often used to evaluate forecasting techniques. Fig.6 shows examples of these diagrams.

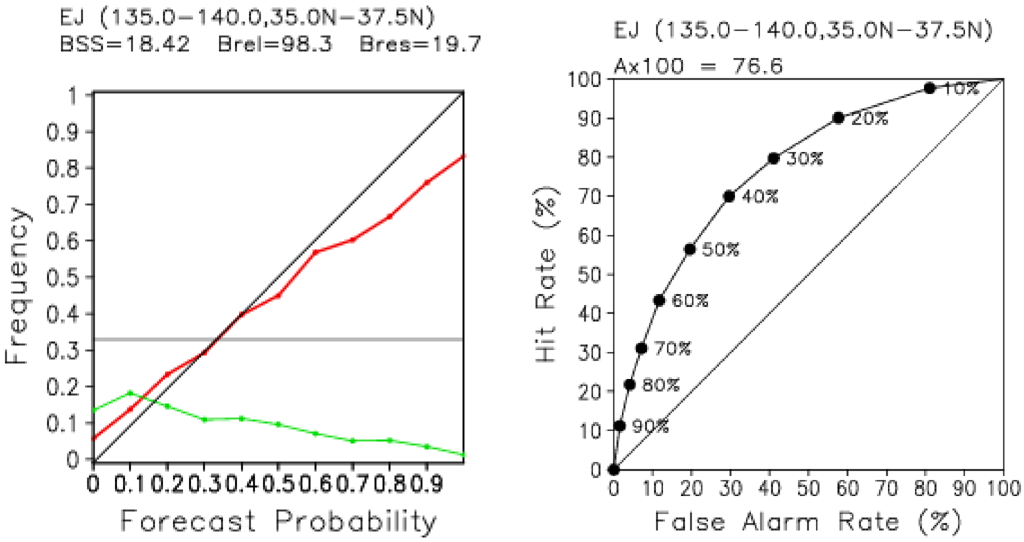


Figure 6 Performance of guidance for monthly forecasts for Eastern Japan
Left: Reliability diagram; Right: ROC curve
The red line in the diagram on the left shows occurrence frequencies by forecast probability, and the green line shows frequencies of forecast probabilities.
The curve in the diagram on the right shows combinations of hit rates and false alarm rates by forecast probability value.

a) Brier Score

The Brier score (hereinafter abbreviated to BS) is used as a score to enable a comprehensive evaluation of probability forecasting. The BS for two categories (written as b) is defined by formula (1):

$$b = \frac{1}{N}\sum_{i=1}^{N}(p_i - v_i), \quad 0 \le p_i \le 1, \quad v_i \in \{0,1\} \tag{1}$$

where N represents the total number of forecasts, p_i a forecast probability value, and v_i a variable that assumes 1 when the predicted phenomenon occurs and 0 when it does not occur. A smaller value of b represents a better result, and the BS is 0 for perfect prediction, where forecasts are deterministic, with probability values assuming either 0 or 100%, and where all of the forecasts come true.

Murphy (1973) showed that b can be mathematically decomposed into three terms. If we rewrite (1) by separating terms for occurrence of the phenomenon from those for non-occurrence, we obtain

$$b = \frac{1}{N}\left\{\sum_t (p_t - 1)^2 M_t + \sum_t (p_t - 0)^2 (N_t - M_t)\right\} \tag{2}$$

where t is a subscript assigned to each probability value (for example, $p_1 = 0.1$), M_t the number of predicted phenomena that actually occurred for the tth probability value, and N_t the number of forecasts made for the tth probability value. We can transform (2) further to obtain (3):

$$b = \sum_t (p_t - \frac{M_t}{N_t})^2 \frac{N_t}{N} - \sum_t (\frac{M}{N} - \frac{M_t}{N_t})^2 \frac{N_t}{N} + \frac{M}{N}(1 - \frac{M}{N}) \tag{3}$$

where $M = \sum_t M_t$, which represents the total number of predicted phenomena that actually occurred.

The first term of the right hand side of (3) is called "reliability." If we allow $brel$ to be the reliability, then

$$brel = \sum_t (p_t - \frac{M_t}{N_t})^2 \frac{N_t}{N} \tag{4}$$

where $\frac{M_t}{N_t}$ is the proportion of predicted phenomena that occurred to the total number of forecasts for the tth probability value (hereinafter referred to as the occurrence frequency). A curve obtained by plotting the occurrence frequency for each forecast probability value is known as a reliability diagram. A smaller values of $brel$ represents a larger number of forecasts made with forecast probabilities that match actual occurrence frequencies. If all points of the reliability curve are located on the 45-degree line where forecast probabilities and occurrence frequencies match, $brel = 0$.

The second term of (3) is called "resolution." If we allow it to be represented by $bres$, then

$$bres = \sum_t (\frac{M}{N} - \frac{M_t}{N_t})^2 \frac{N_t}{N} \tag{5}$$

The negative sign of $bres$ in (3) indicates that larger values of $bres$ represent better results. A larger value of $bres$ indicates that the difference between the occurrence frequency of each forecast probability and the climatic frequency of the forecast event $\frac{M}{N}$ is greater. An examination of the reliability curve reveals that since the climatic frequency $\frac{M}{N}$ for a three-category forecast is 33%, the greater the occurrence frequency of probabilities of 0% or 60% or higher, the greater the value of $bres$ is and the better the results will be.

The third term in (3) is called "uncertainty." If we allow it to be represented by $bunc$, then

$$bunc = \frac{M}{N}(1 - \frac{M}{N}) \qquad (6)$$

This term is not related to the performance of forecasting. It is a term that depends on the climatic frequency alone, which shows the level of uncertainty of the phenomenon.

These scores allow us to make judgments about the relative performance of forecasts, but not about their significance. An improvement rate with respect to a climatic value forecast is often used as a measure of the significance of forecasts. If we write the BS of a climatic value as bc, the improvement rate can be defined by the following formula:

$$BSS = \frac{bc - b}{bc} \qquad (7)$$

BSS is called the "skill score." If $BSS \le 0$, the forecast is inferior to the climatic value forecast, and BSS assumes the maximum value ($BSS = 1$) for a perfect forecast.
Likewise, the improvements rate with respect to a climatic value forecast can be defined for (4) and (5) as follows:

$$Brel = \frac{bc - brel}{bc} \qquad (8)$$

$$Bres = \frac{bres}{bunc} \qquad (9)$$

$Brel = Bres = 1$ for a perfect forecast.

These arguments make it clear that the reliability curve in the diagram on the left in Fig.3, which roughly follows the 45-degree line, shows a high level of reliability in the forecasting. Meanwhile, the occurrence rate assumes the highest value for the forecast probability of 10%, decreasing as the probability value becomes greater. The diagram also shows a positive value for BSS, which indicates that the forecasting is made with a higher level of skill compared to the forecast of climatic values.

b) ROC

An ROC (relative operating characteristic) curve is a diagram used to verify and evaluate probability forecasting based on economic values (user cost or loss). The curve is drawn to evaluate forecasting, using "hit rates" on the vertical axis and "false alarm rates" on the horizontal axis.
Table 1 shows the relationship between forecasts and observations for a case where an

event is predicted to occur with forecast probabilities of p_i or higher. The "hit rate" hr_i and the "false alarm rate" fr_i are defined as follows:

$$hr_i = \frac{A_i}{A_i + C_i} = \frac{\sum_{t=i}^{10} M_t}{M} \qquad (10)$$

$$fr_i = \frac{B_i}{B_i + D_i} = \frac{\sum_{t=i}^{10} (N_t - M_t)}{N - M} \qquad (11)$$

hr_i represents the proportion of predicted phenomena out of the total number of phenomena that actually occurred, and is called a "hit rate." The proportion of predicted phenomena that occurred to the total number of forecasts is often called a hit rate. The hit rate for an ROC curve refers to the "level of coverage," which represents the degree to which phenomena are covered by forecasts. Meanwhile, fr_i represents the proportion of predicted phenomena out of the total number of phenomena that did not occur, and is called a "false alarm rate."

		Observation	
		Yes	No
Forecast	Yes	Ai	Bi
	No	Ci	Di

Table 1 Correspondence between forecast and observation for a case where a phenomenon was predicted with a forecast probability of p_i or higher

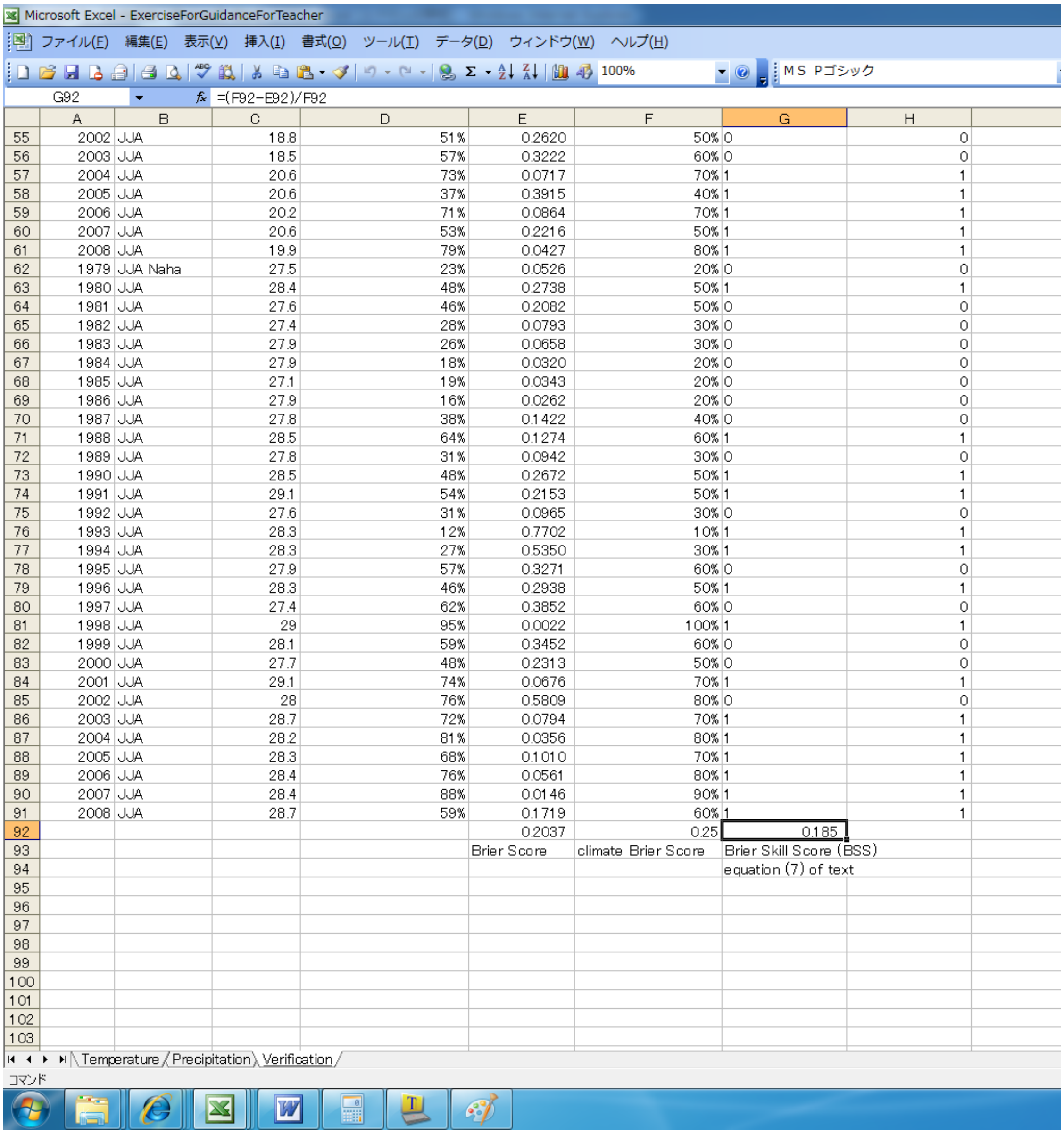
Exercise

A lot of case is needed to verify probabilistic forecast, so try to make some guidance at weather stations. It is possible to total all case of each station and season. But it is not necessary to total different variables such as temperature and precipitation. Because the skill generally varies with variables.

If you can get them enough, let's try to make a reliability diagram and calculate Brier Skill Score using Excel software.

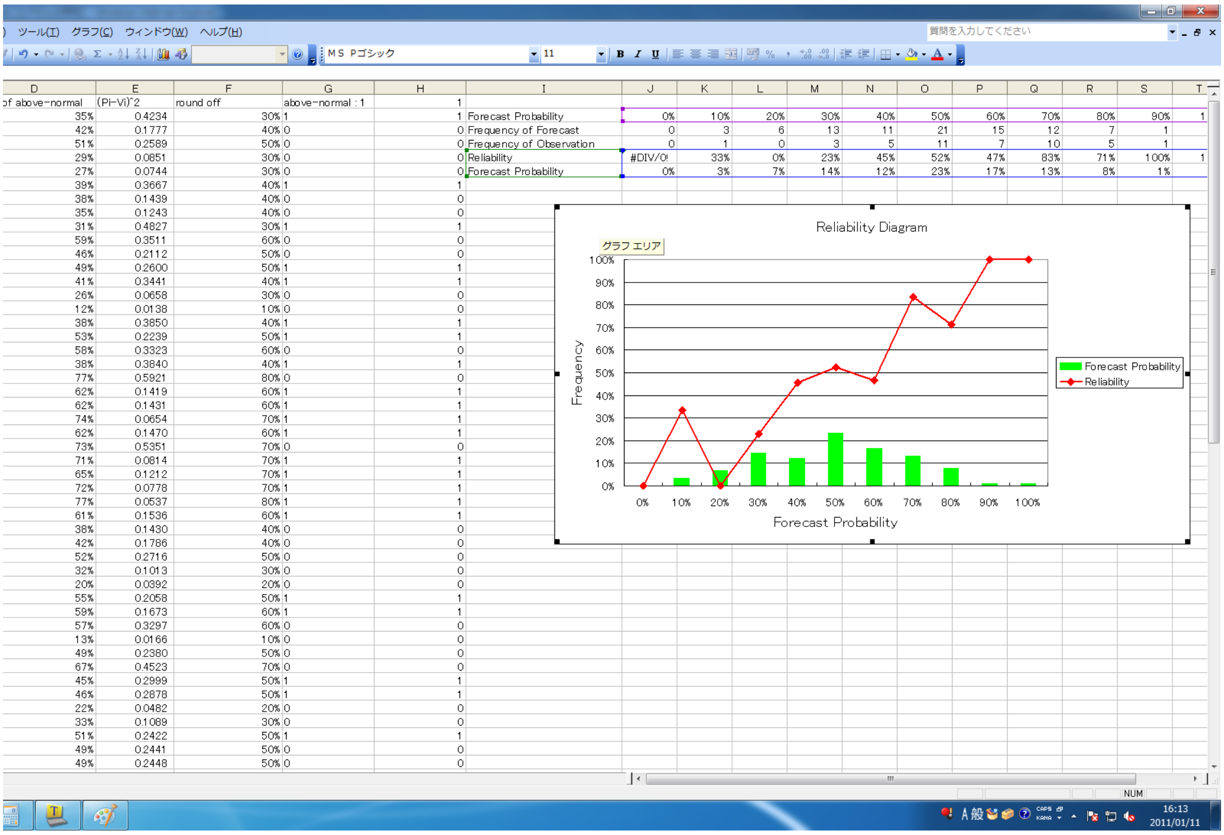
Question

Can you get more accuracy guidance than the following Japanese guidance?



	A	B	C	D	E	F	G	H
55	2002	JJA	18.8	51%	0.2620	50%	0	0
56	2003	JJA	18.5	57%	0.3222	60%	0	0
57	2004	JJA	20.6	73%	0.0717	70%	1	1
58	2005	JJA	20.6	37%	0.3915	40%	1	1
59	2006	JJA	20.2	71%	0.0864	70%	1	1
60	2007	JJA	20.6	53%	0.2216	50%	1	1
61	2008	JJA	19.9	79%	0.0427	80%	1	1
62	1979	JJA Naha	27.5	23%	0.0526	20%	0	0
63	1980	JJA	28.4	48%	0.2738	50%	1	1
64	1981	JJA	27.6	46%	0.2082	50%	0	0
65	1982	JJA	27.4	28%	0.0793	30%	0	0
66	1983	JJA	27.9	26%	0.0658	30%	0	0
67	1984	JJA	27.9	18%	0.0320	20%	0	0
68	1985	JJA	27.1	19%	0.0343	20%	0	0
69	1986	JJA	27.9	16%	0.0262	20%	0	0
70	1987	JJA	27.8	38%	0.1422	40%	0	0
71	1988	JJA	28.5	64%	0.1274	60%	1	1
72	1989	JJA	27.8	31%	0.0942	30%	0	0
73	1990	JJA	28.5	48%	0.2672	50%	1	1
74	1991	JJA	29.1	54%	0.2153	50%	1	1
75	1992	JJA	27.6	31%	0.0965	30%	0	0
76	1993	JJA	28.3	12%	0.7702	10%	1	1
77	1994	JJA	28.3	27%	0.5350	30%	1	1
78	1995	JJA	27.9	57%	0.3271	60%	0	0
79	1996	JJA	28.3	46%	0.2938	50%	1	1
80	1997	JJA	27.4	62%	0.3852	60%	0	0
81	1998	JJA	29	95%	0.0022	100%	1	1
82	1999	JJA	28.1	59%	0.3452	60%	0	0
83	2000	JJA	27.7	48%	0.2313	50%	0	0
84	2001	JJA	29.1	74%	0.0676	70%	1	1
85	2002	JJA	28	76%	0.5809	80%	0	0
86	2003	JJA	28.7	72%	0.0794	70%	1	1
87	2004	JJA	28.2	81%	0.0356	80%	1	1
88	2005	JJA	28.3	68%	0.1010	70%	1	1
89	2006	JJA	28.4	76%	0.0561	80%	1	1
90	2007	JJA	28.4	88%	0.0146	90%	1	1
91	2008	JJA	28.7	59%	0.1719	60%	1	1
92					0.2037	0.25	0.185	
93					Brier Score	climate Brier Score	Brier Skill Score (BSS)	
94							equation (7) of text	
95								
96								
97								
98								
99								
100								
101								
102								
103								

For Example: Calculating Brier Skill Score of JJA temperature, totaling Tokyo, Sapporo and Naha station in Japan.



For Example: Drawing Reliability Diagram of JJA temperature, totaling Tokyo, Sapporo and Naha station in Japan.

Reference

Murphy, A. H. 1977: The value of climatorogical categories and probabilistic forecast in the cost-loss ratio situations. . Mon. Weather Rev. 105 803-816

Palmer, T. N. ,C. Brankovic and D.S. Richardson 2000: A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations.