Exercise Part 1
- Producing Guidance and Verification -

TCC Training Seminar on Seasonal Prediction Products

11-15 November 2013

Objectives

- To understand how to produce guidance for seasonal temperature and precipitation.
- To think rationally and find possible predictors.
- To find more effective predictors.
- To verify the guidance based on deterministic and probabilistic methods.

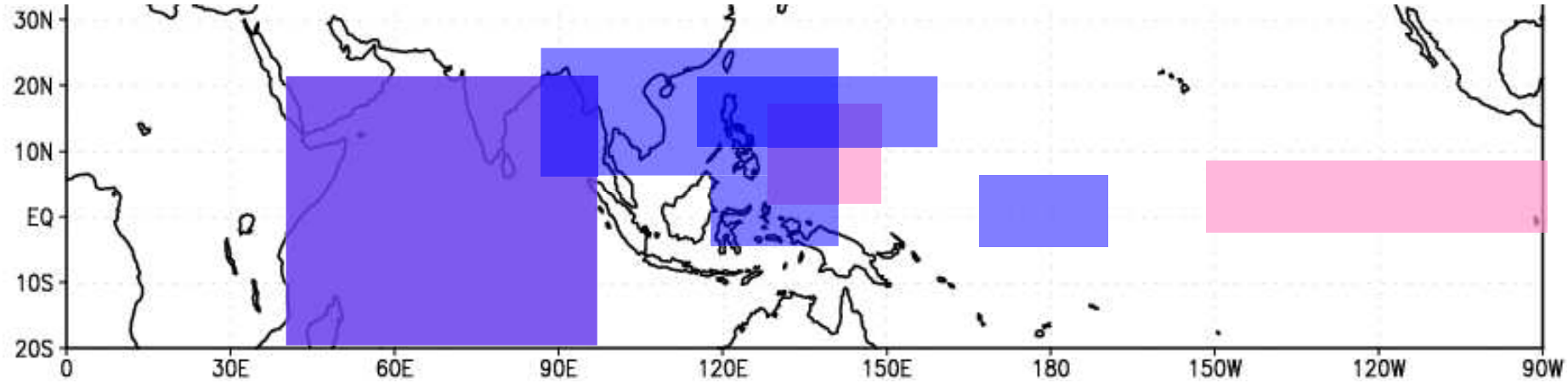
Contents

Preparations

- Observation data (You have prepared.)
- Predictors (init. 28th Sep. for DJF)
  - GPV data over your stations "*GPVdata.xls*"
  - Indices such as NINO3SST, Asia Monsoon, etc. "*indices.xls*"
- Excel software for making guidance "*ProducingGuidance.xls*"
- Excel software for verification "*Verification.xls*"

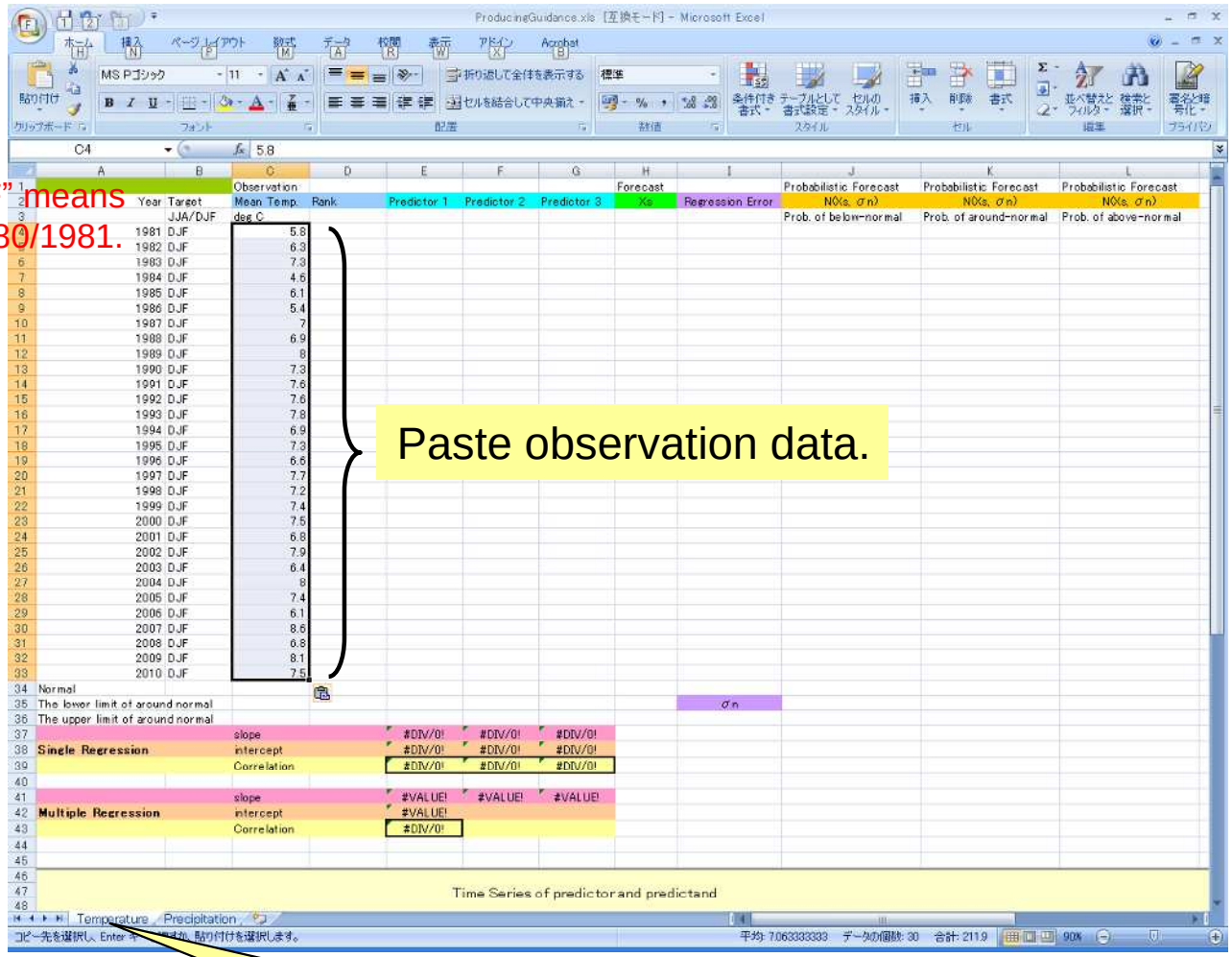
Indices

	indices	variables	areas
SST	NINO3 SST	SST	(150W-90W, 5S-5N)
	NINOWEST SST	SST	(130E-150E, EQ-15N)
	IOBW SST	SST	(40E-100E, 20S-20N)
	WIO SST	SST	(40E-70E, 0-20N)
	EIO SST	SST	(70E-100E, 0-20N)
RAIN	IOBW RAIN	RAIN	(40E-100E, 20S-20N)
	WIO RAIN	RAIN	(40E-70E, 0-20N)
	EIO RAIN	RAIN	(70E-100E, 0-20N)
	SAMOI RAIN	RAIN	(80E-140E, 5N-25N)
	WNP RAIN	RAIN	(110E-160E, 10N-20N)
	SEAsia RAIN	RAIN	(115E-140E, 10N-20N)
	MC RAIN	RAIN	(110E-135E, 5S-5N)
	DL RAIN	RAIN	(170E-170W, 5S-5N)
Z500	Z2030	500hPa Height	(0-360, 20N-30N)
	Z3040	500hPa Height	(0-360, 30N-40N)
	Z4050	500hPa Height	(0-360, 40N-50N)
	Z5060	500hPa Height	(0-360, 50N-60N)
Thickness	THMD	Thickness Middle	(0-360, 30N-50N, 300hPa-850hPa)
	THEX	Thickness extratropic	(0-360, 30N-90N, 300hPa-850hPa)
	THTR	Thickness tropic	(0-360, 25S-25N, 100hPa-850hPa)



1. Single Regression Model

- Open the ProducingGuidance.xls.
- Paste observation data into the Temperature/Precipitation worksheet.



"1981 DJF" means DJF of 1980/1981.

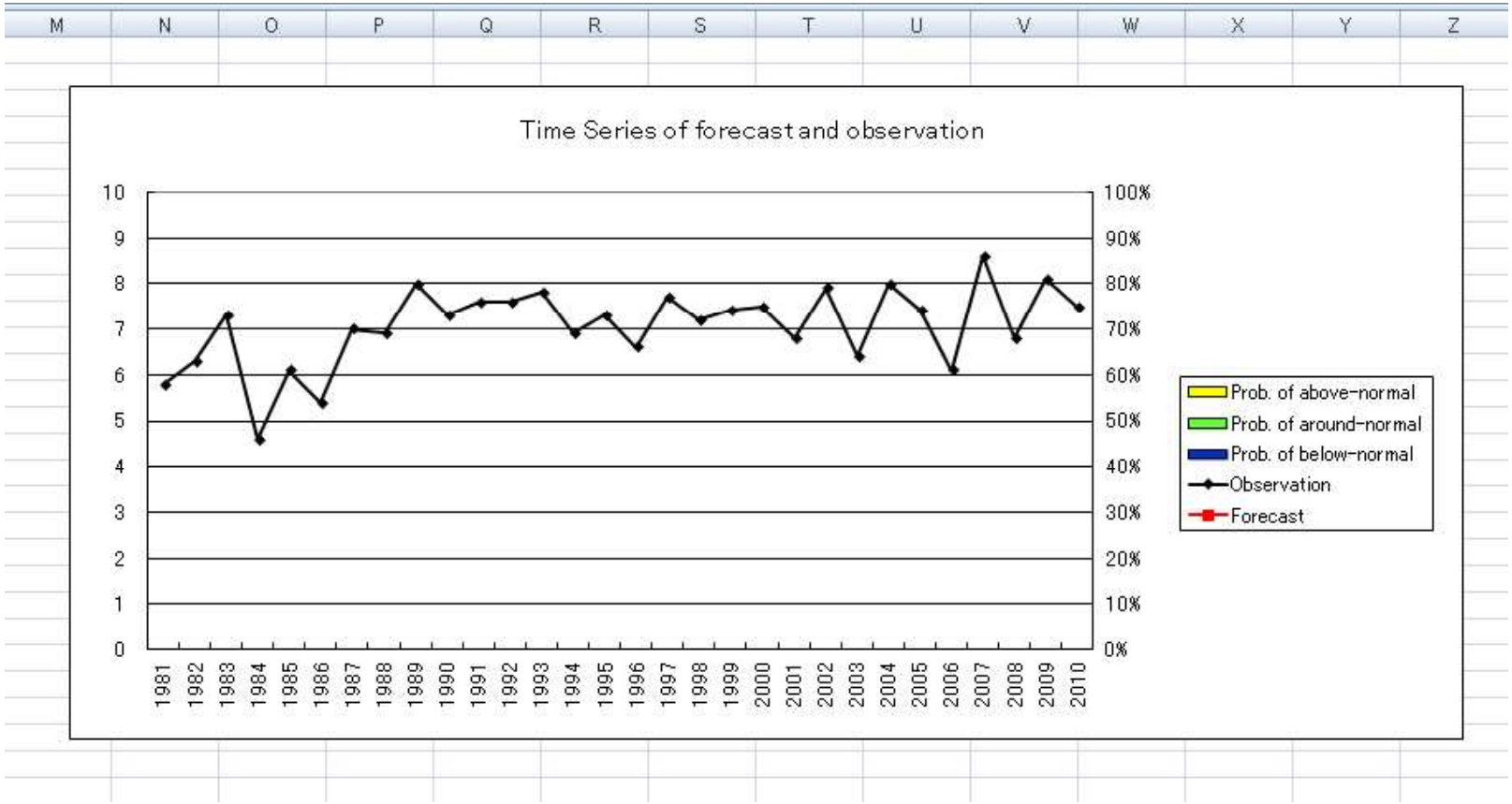
Paste observation data.

Temperature worksheet

In this case, the predictand is temperature and the station is Tokyo.

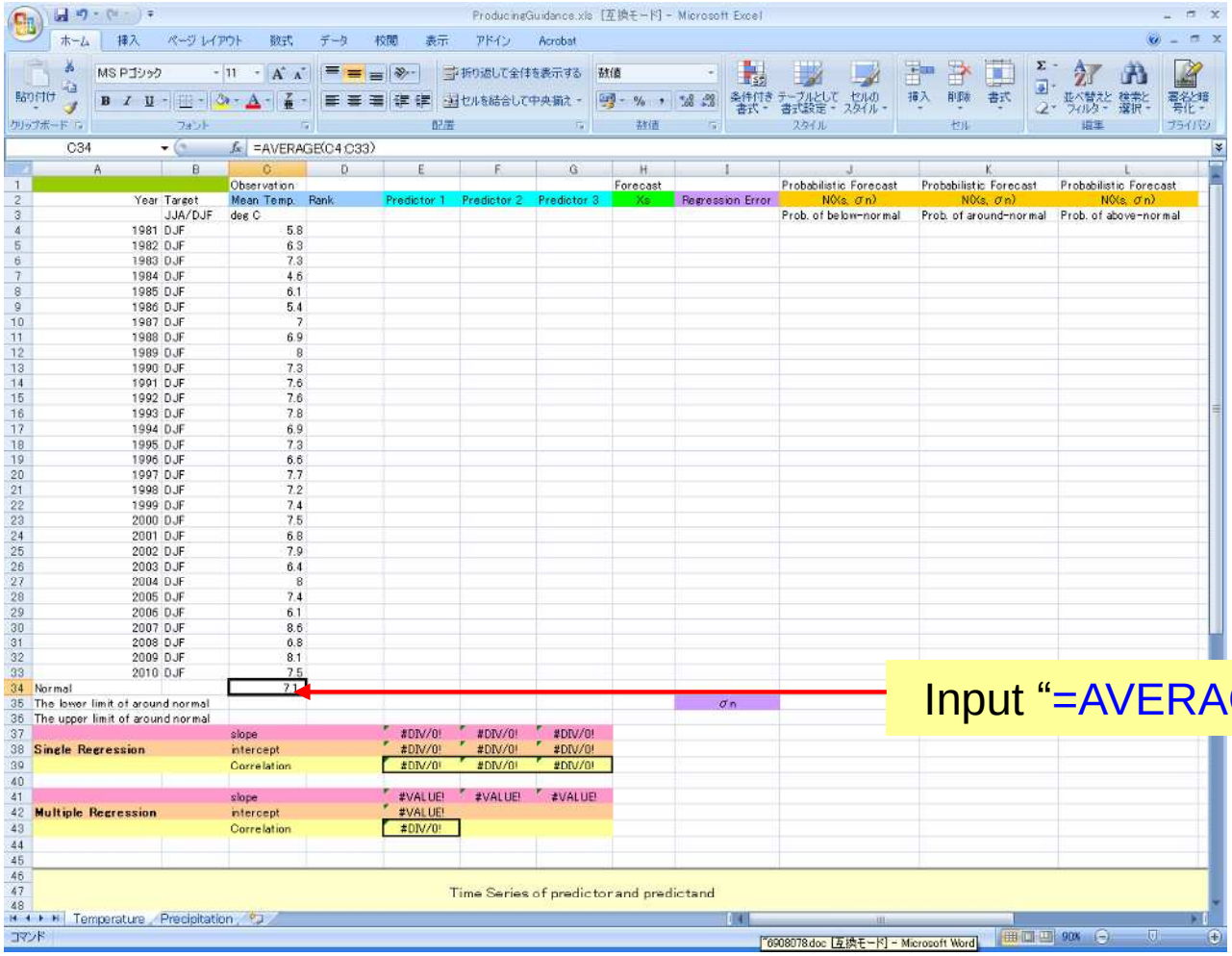
1. Single Regression Model

A graph of time series of the observation data has been automatically produced.



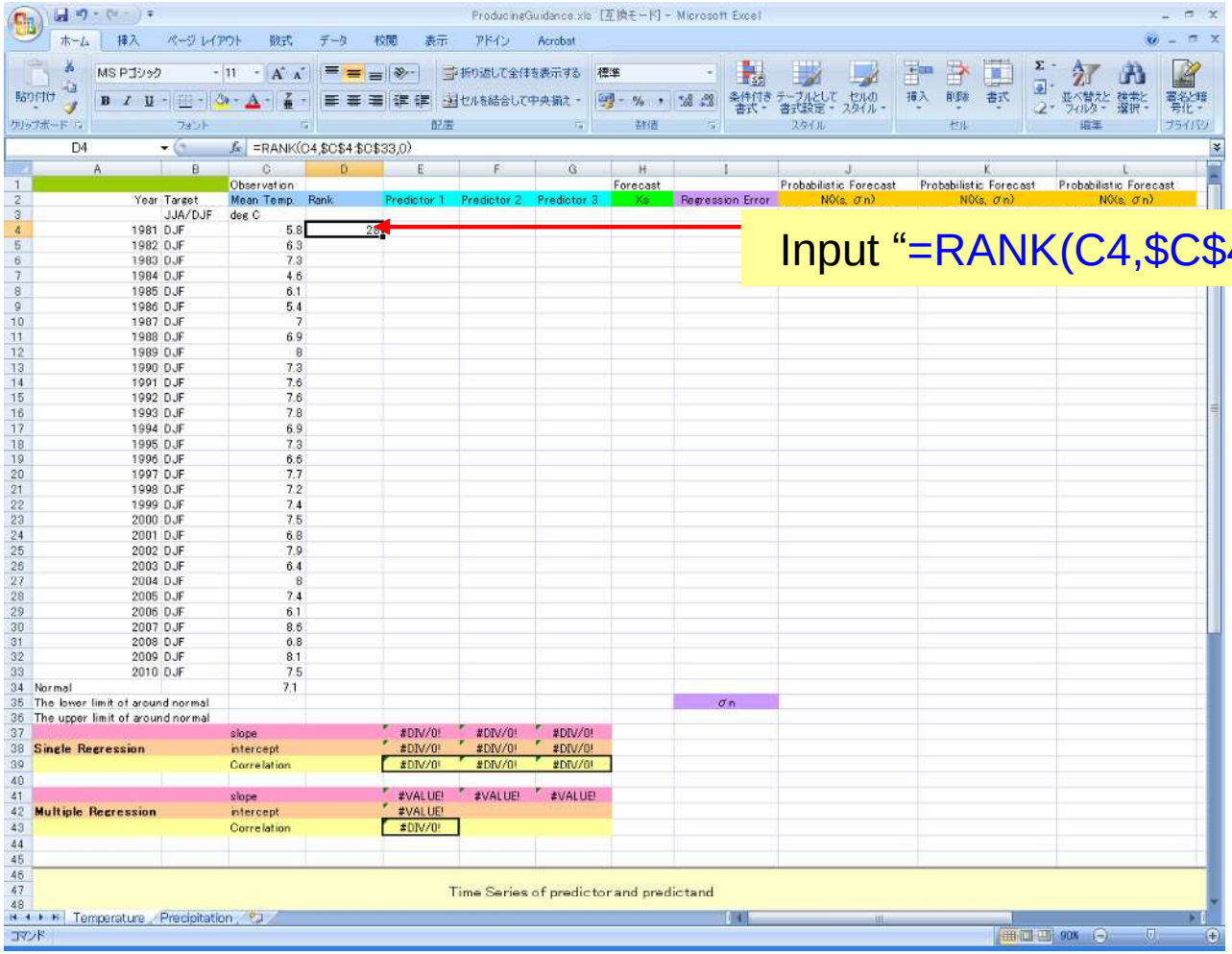
1. Single Regression Model

- Calculate the normal from 1981 to 2010.



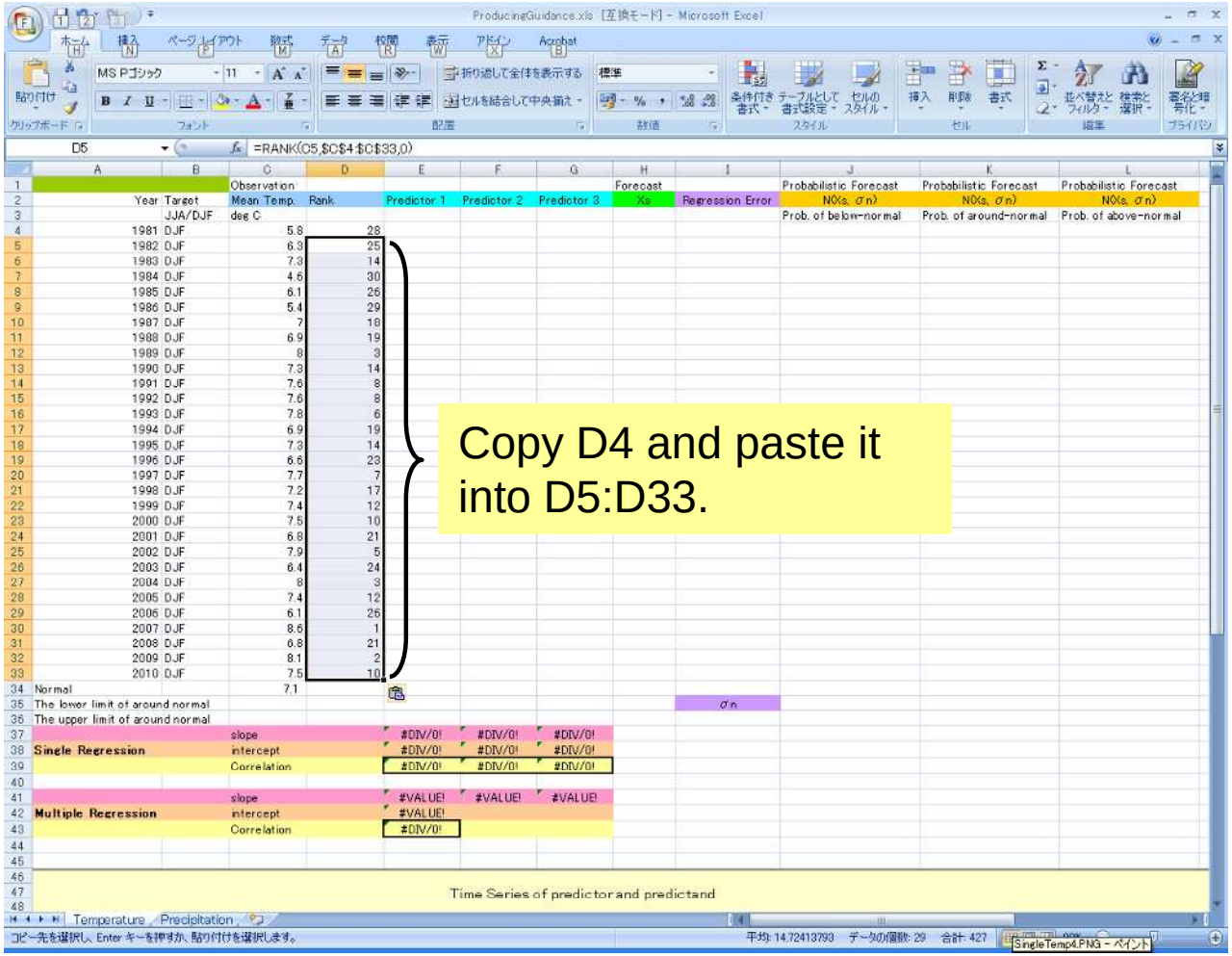
1. Single Regression Model

- Rank the data from 1981 to 2010 in descending order (preparation for probabilistic forecast).



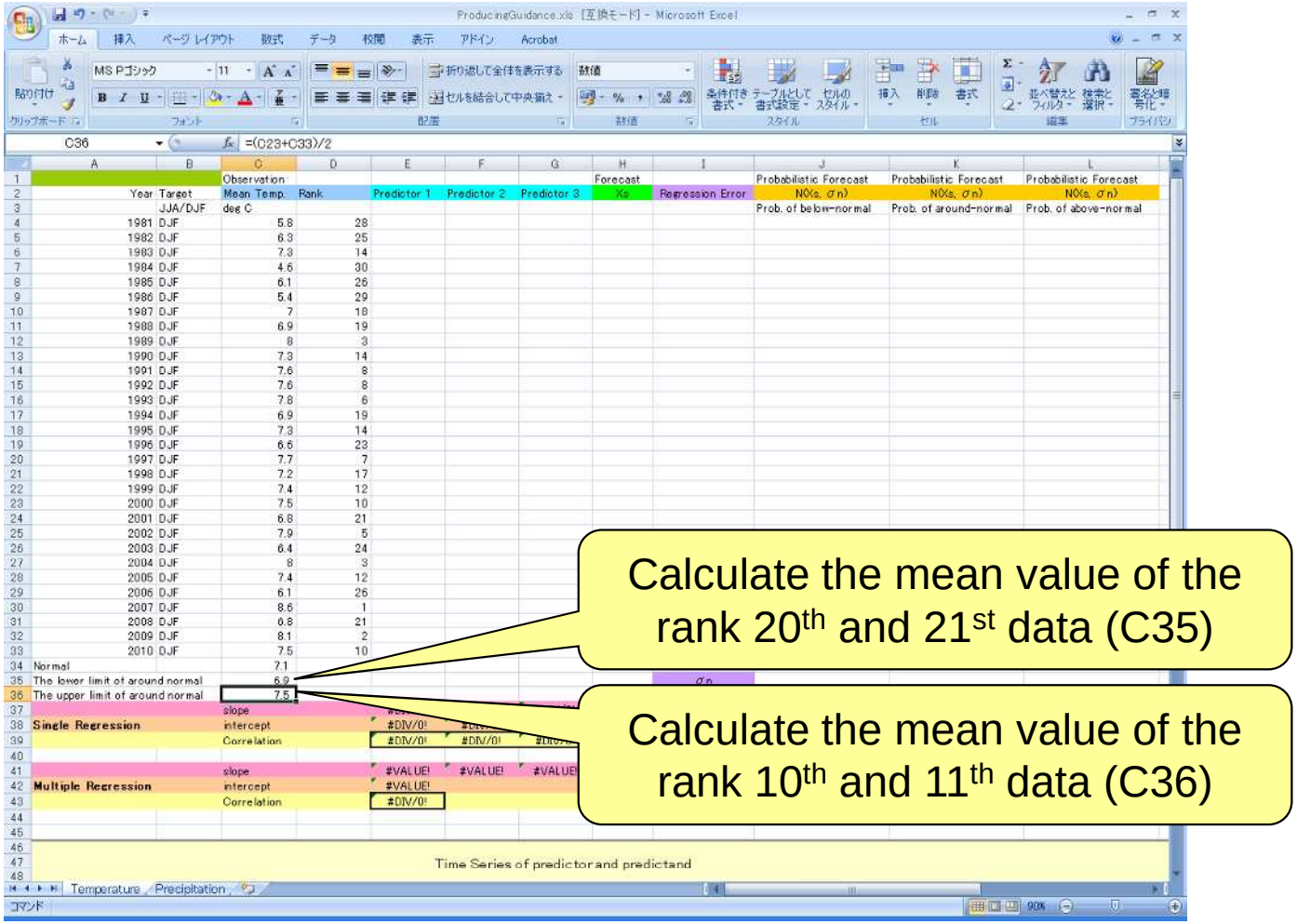
1. Single Regression Model

- Rank the data from 1981 to 2010 (preparation for probabilistic forecast).



1. Single Regression Model

- Calculate the upper and lower limits of around normal (preparation for probabilistic forecast).



1. Single Regression Model

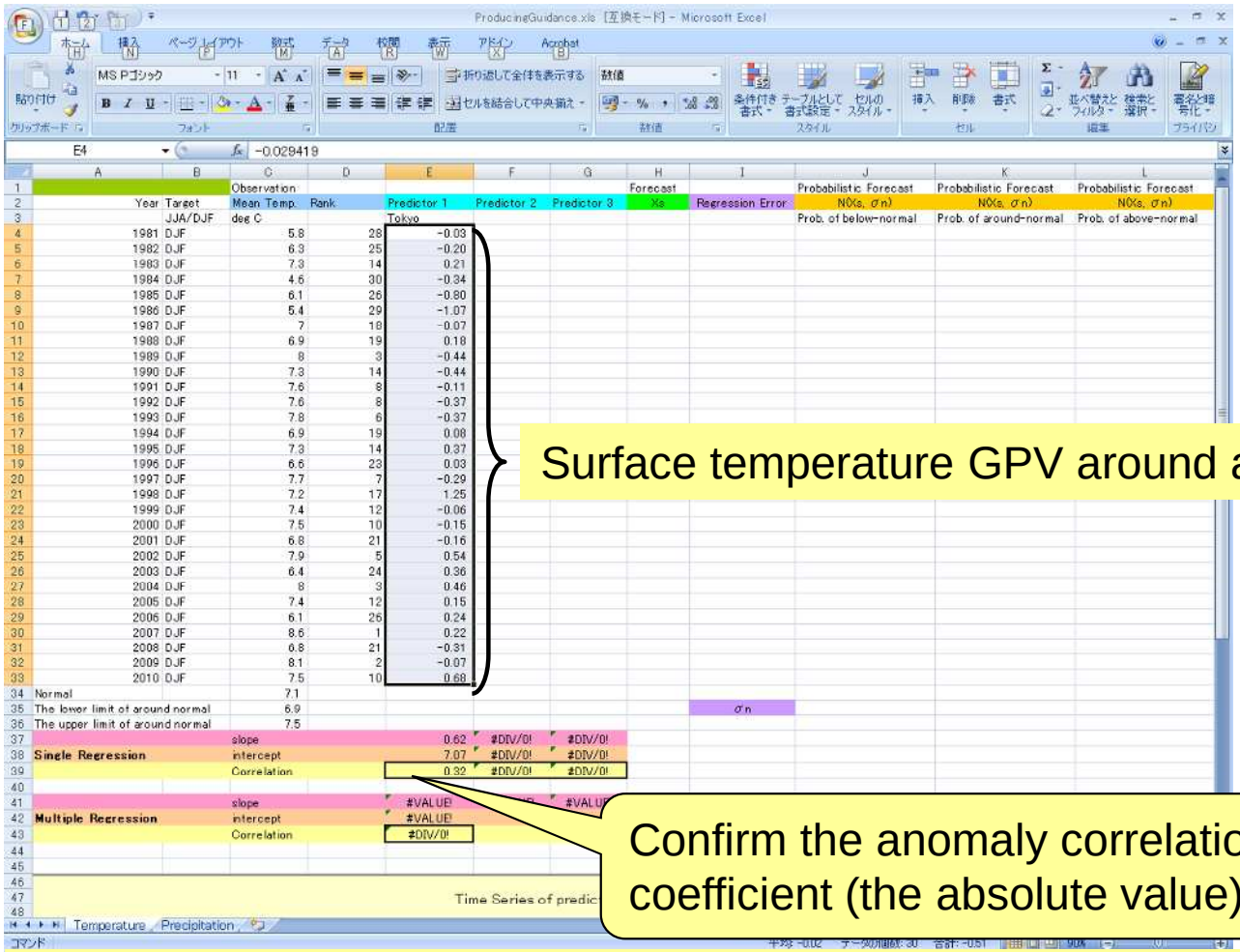
Original line number	Year	Mean Temp	Rank
30	2007	8.6	1
32	2009	8.1	2
12	1989	8	3
27	2004	8	3
25	2002	7.9	5
16	1993	7.8	6
20	1997	7.7	7
14	1991	7.6	8
15	1992	7.6	8
23	2000	7.5	10
33	2010	7.5	10
22	1999	7.4	12
28	2005	7.4	12
6	1983	7.3	14
13	1990	7.3	14
18	1995	7.3	14
21	1998	7.2	17
10	1987	7	18
11	1988	6.9	19
17	1994	6.9	19
24	2001	6.8	21
31	2008	6.8	21
19	1996	6.6	23
26	2003	6.4	24
5	1982	6.3	25
8	1985	6.1	26
29	2006	6.1	26
4	1981	5.8	28
9	1986	5.4	29
7	1984	4.6	30

The upper limit of around normal can be calculated as the average of “C23” and “C33”.

The lower limit of around normal can be calculated as the average of “C17”(“C11) and “C24” (“C31)”.

1. Single Regression Model

- Open GPVdata.xls and indices.xls. ← Predictors
- Select a predictor and paste the data into E column.



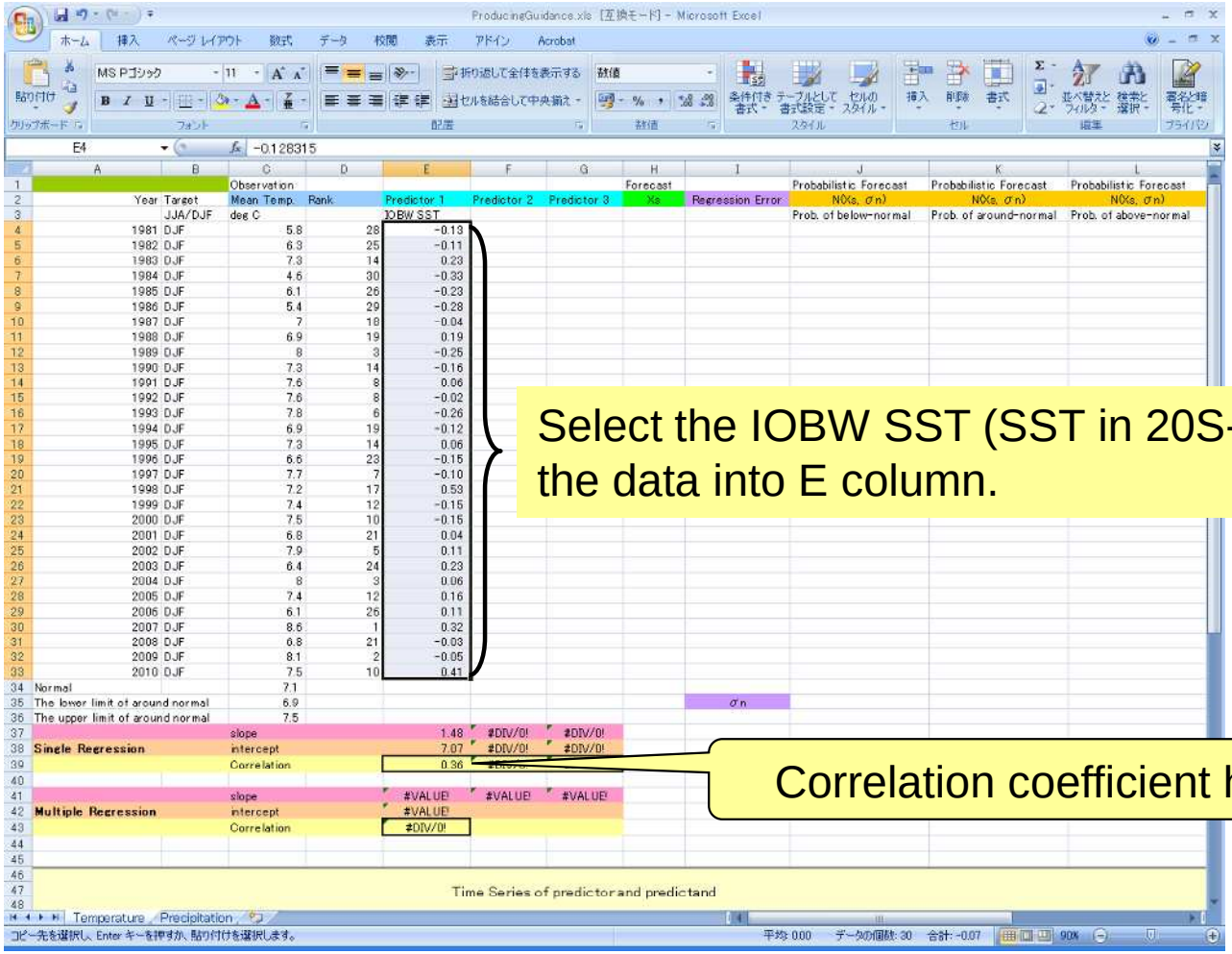
Surface temperature GPV around at the station as a predictor.

Confirm the anomaly correlation coefficient (the absolute value) at E39.

In case of Tokyo, the anomaly correlation is 0.32 (not so large).
I will try other predictors.

1. Single Regression Model

- Try other predictors.



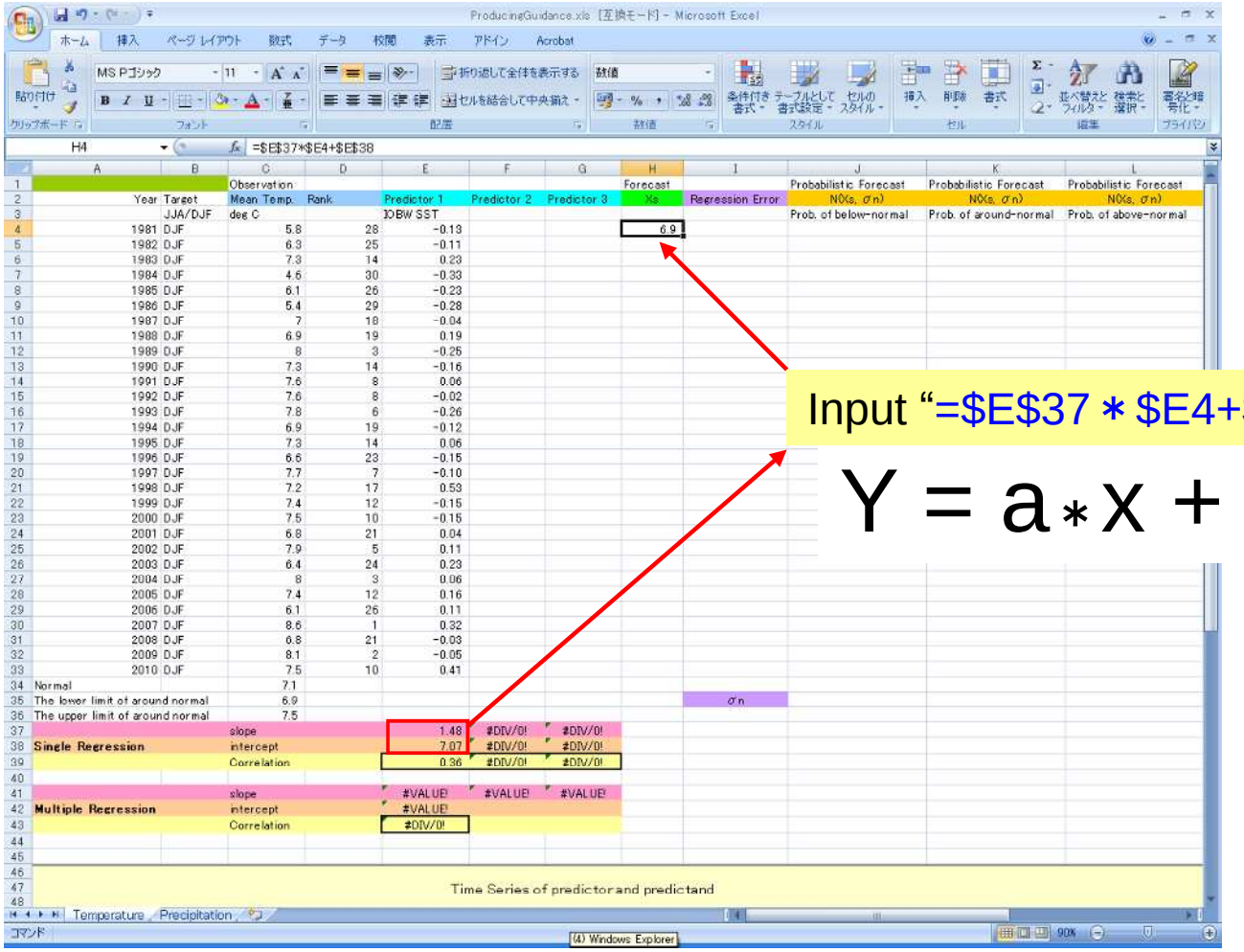
Select the IOBW SST (SST in 20S-20N, 40-100E) and paste the data into E column.

Correlation coefficient has increased to 0.36.

Please try to look for the most effective predictor.

1. Single Regression Model

- Calculate the forecasts using single regression equation.

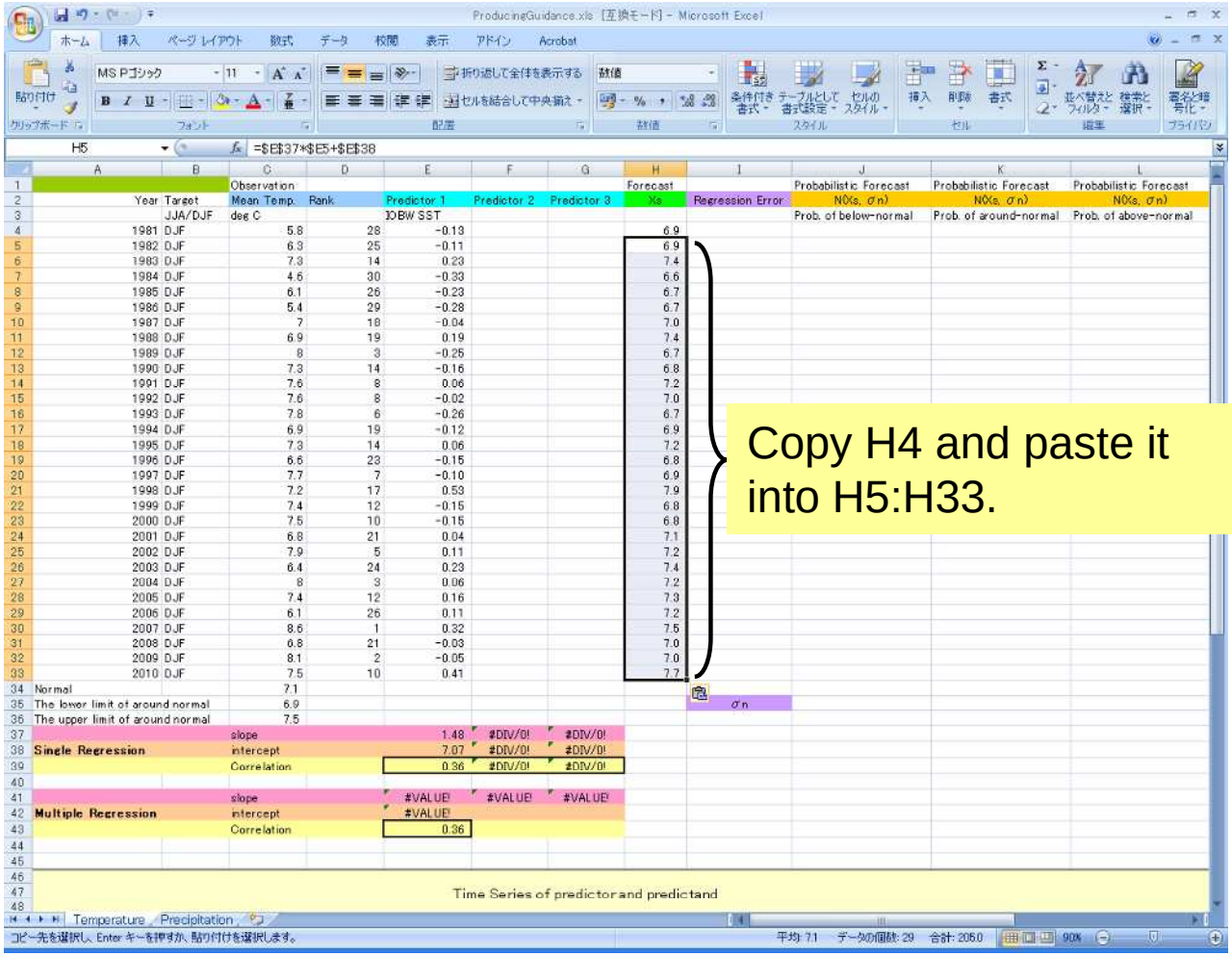


Input “=$E$37 * $E4+$E$38” at H4.

$$Y = a * x + b$$

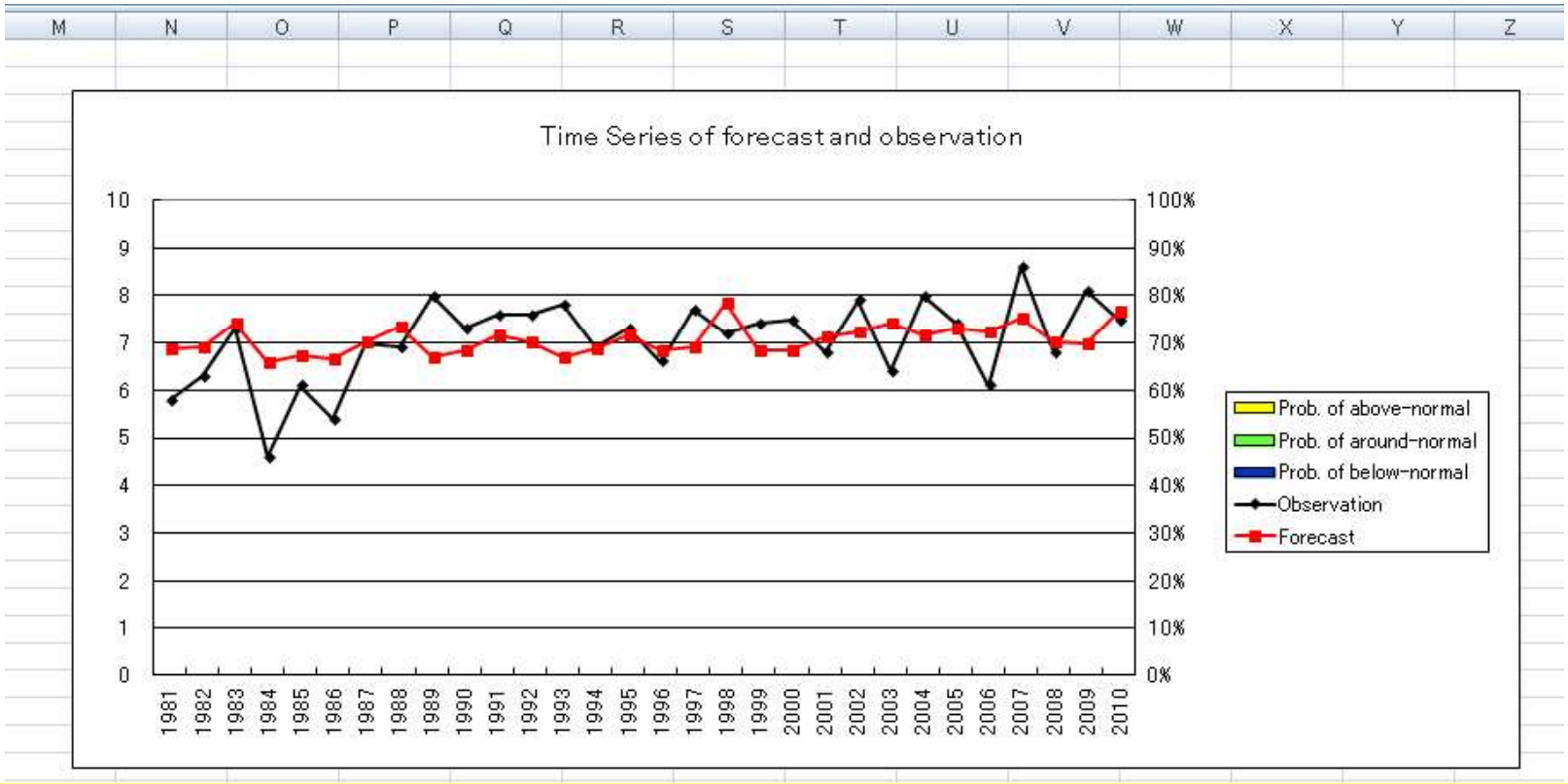
1. Single Regression Model

- Calculate the forecasts using single regression equation.



1. Single Regression Model

The time series of the forecast data has been automatically added to the graph.

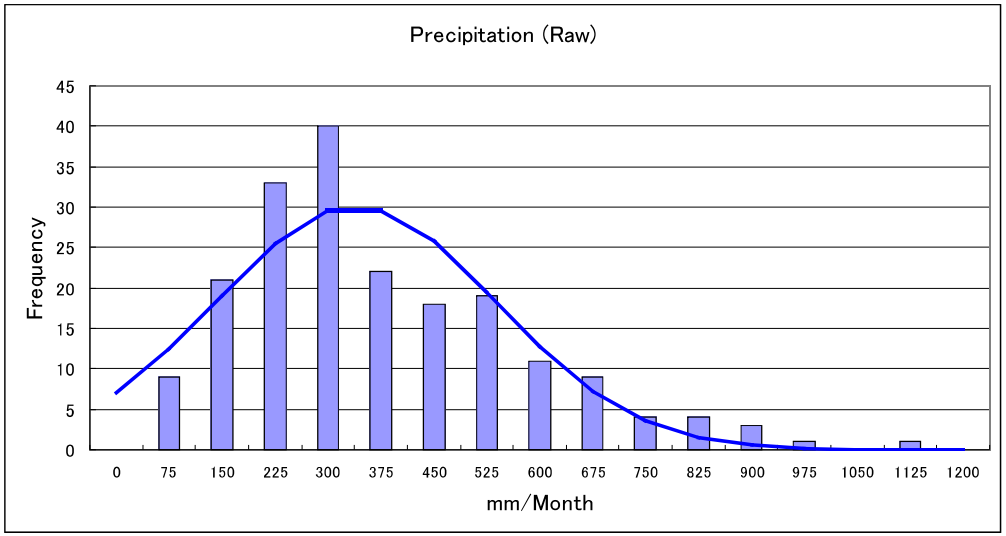


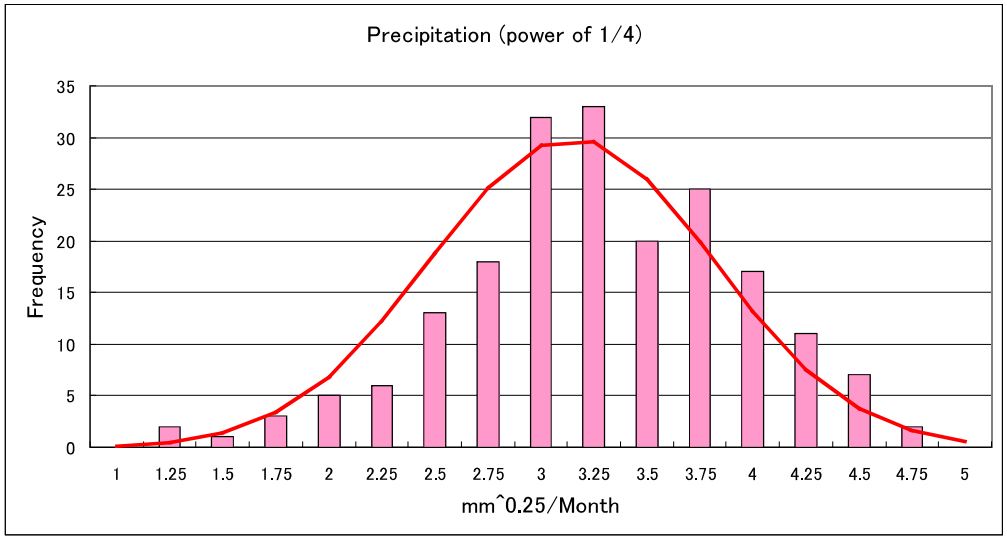
I will ask you the following questions at the end of this session:
1. “What predictor did you select?”
2. “Could you get an accurate guidance?”

1. Single Regression Model

How about 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 normalization is necessary for precipitation. JMA seasonal forecast guidance uses the 1/4-power transformation as the simplest method.



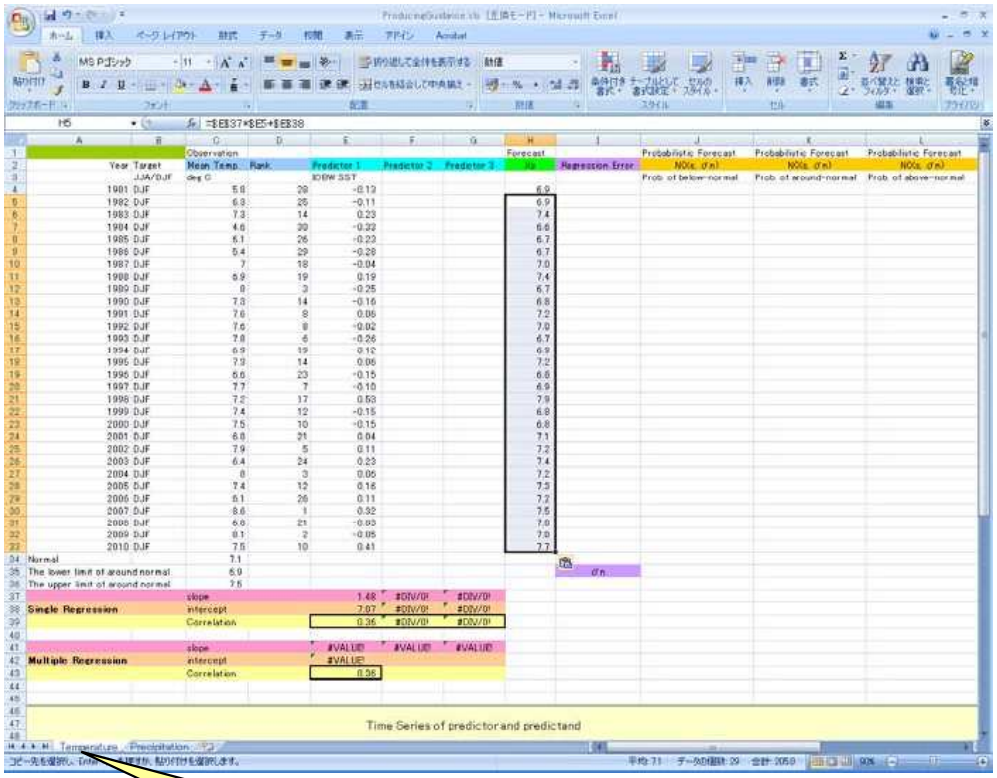


The histogram of raw precipitation data has gaps from Gaussian distribution (left).
The histogram of the power of 1/4 of precipitation values is similar to Gaussian distribution (right).

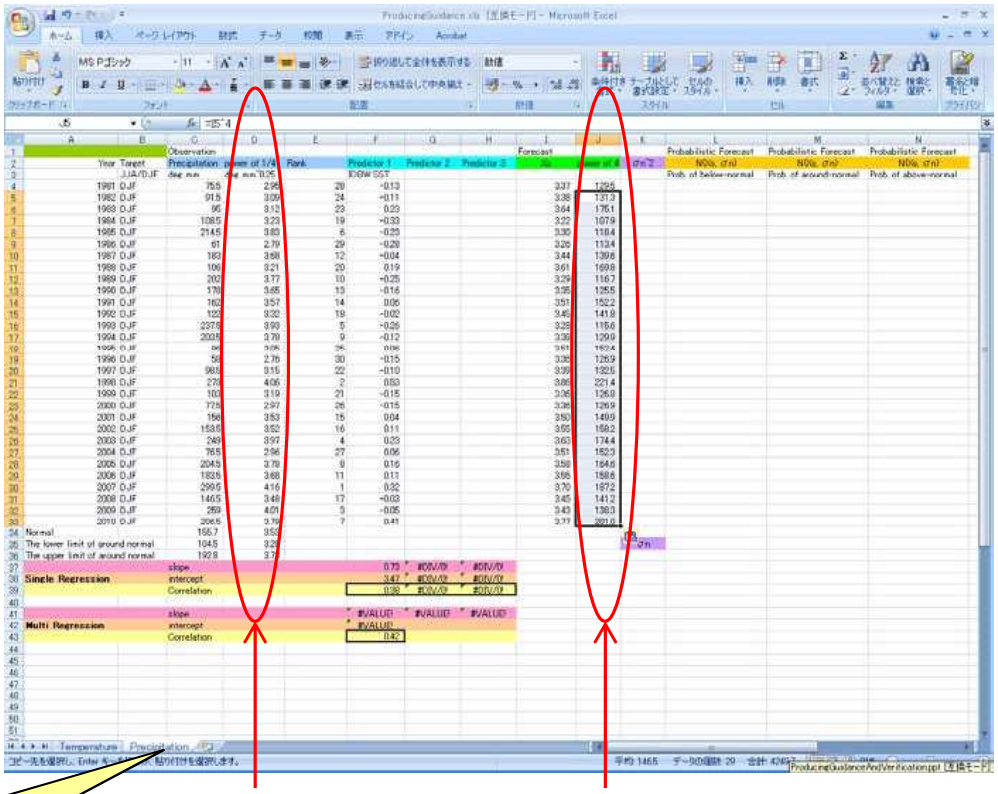
1. Single Regression Model

How about precipitation?

Temperature worksheet



Precipitation worksheet



Temperature worksheet

Precipitation worksheet

the 1/4-power transformation

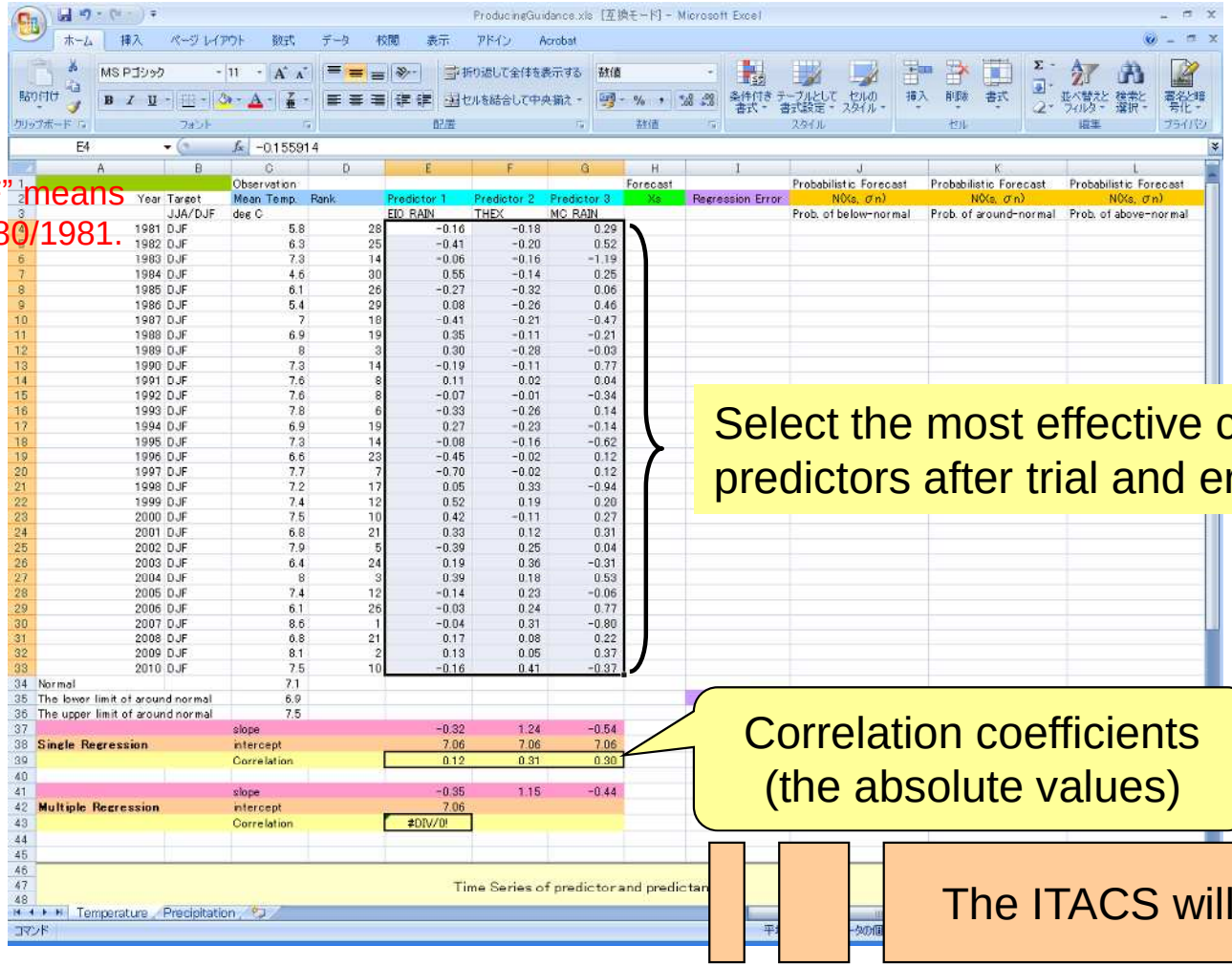
the 4th-power transformation

Precipitation worksheet has two additional columns related with “normalization”.
Please see the appendix slides to know the detail.

2. Multiple Regression Model

- Try to look for most effective combination of predictors.

"1981 DJF" means DJF of 1980/1981.



	A	B	C	D	E	F	G	H	I	J	K	L
1			Observation					Forecast		Probabilistic Forecast	Probabilistic Forecast	Probabilistic Forecast
2	Year	Target	Mean Temp.	Rank	Predictor 1	Predictor 2	Predictor 3	Xa	Regression Error	N(Xa, σ'n)	N(Xa, σ'n)	N(Xa, σ'n)
3		JJA/DJF	deg C		EIO_RAIN	THEX	MC_RAIN			Prob. of below-normal	Prob. of around-normal	Prob. of above-normal
4	1981	DJF	5.8	28	-0.16	-0.18	0.29					
5	1982	DJF	6.3	25	-0.41	-0.20	0.52					
6	1983	DJF	7.3	14	-0.06	-0.16	-1.19					
7	1984	DJF	4.6	30	0.55	-0.14	0.25					
8	1985	DJF	6.1	26	-0.27	-0.32	0.06					
9	1986	DJF	5.4	29	0.08	-0.26	0.46					
10	1987	DJF	7	18	-0.41	-0.21	-0.47					
11	1988	DJF	6.9	19	0.35	-0.11	-0.21					
12	1989	DJF	8	3	0.30	-0.28	-0.03					
13	1990	DJF	7.3	14	-0.19	-0.11	0.77					
14	1991	DJF	7.6	8	0.11	0.02	0.04					
15	1992	DJF	7.6	8	-0.07	-0.01	-0.34					
16	1993	DJF	7.8	6	-0.33	-0.26	0.14					
17	1994	DJF	6.9	19	0.27	-0.23	-0.14					
18	1995	DJF	7.3	14	-0.08	-0.16	-0.62					
19	1996	DJF	6.6	23	-0.45	-0.02	0.12					
20	1997	DJF	7.7	7	-0.70	-0.02	0.12					
21	1998	DJF	7.2	17	0.05	0.33	-0.94					
22	1999	DJF	7.4	12	0.52	0.19	0.20					
23	2000	DJF	7.5	10	0.42	-0.11	0.27					
24	2001	DJF	6.8	21	0.33	0.12	0.31					
25	2002	DJF	7.9	5	-0.39	0.25	0.04					
26	2003	DJF	6.4	24	0.19	0.36	-0.31					
27	2004	DJF	8	3	0.39	0.18	0.53					
28	2005	DJF	7.4	12	-0.14	0.23	-0.06					
29	2006	DJF	6.1	26	-0.03	0.24	0.77					
30	2007	DJF	8.6	1	-0.04	0.31	-0.80					
31	2008	DJF	6.8	21	0.17	0.08	0.22					
32	2009	DJF	8.1	2	0.13	0.05	0.37					
33	2010	DJF	7.5	10	-0.16	0.41	-0.37					
34	Normal		7.1									
35	The lower limit of around normal		6.9									
36	The upper limit of around normal		7.5									
37			slope		-0.32	1.24	-0.54					
38	Single Regression		intercept		7.06	7.06	7.06					
39			Correlation		0.12	0.31	0.30					
40												
41			slope		-0.35	1.15	-0.44					
42	Multiple Regression		intercept		7.06							
43			Correlation		#DIV/0!							
44												
45												
46												
47					Time Series of predictor and predictan							
48												

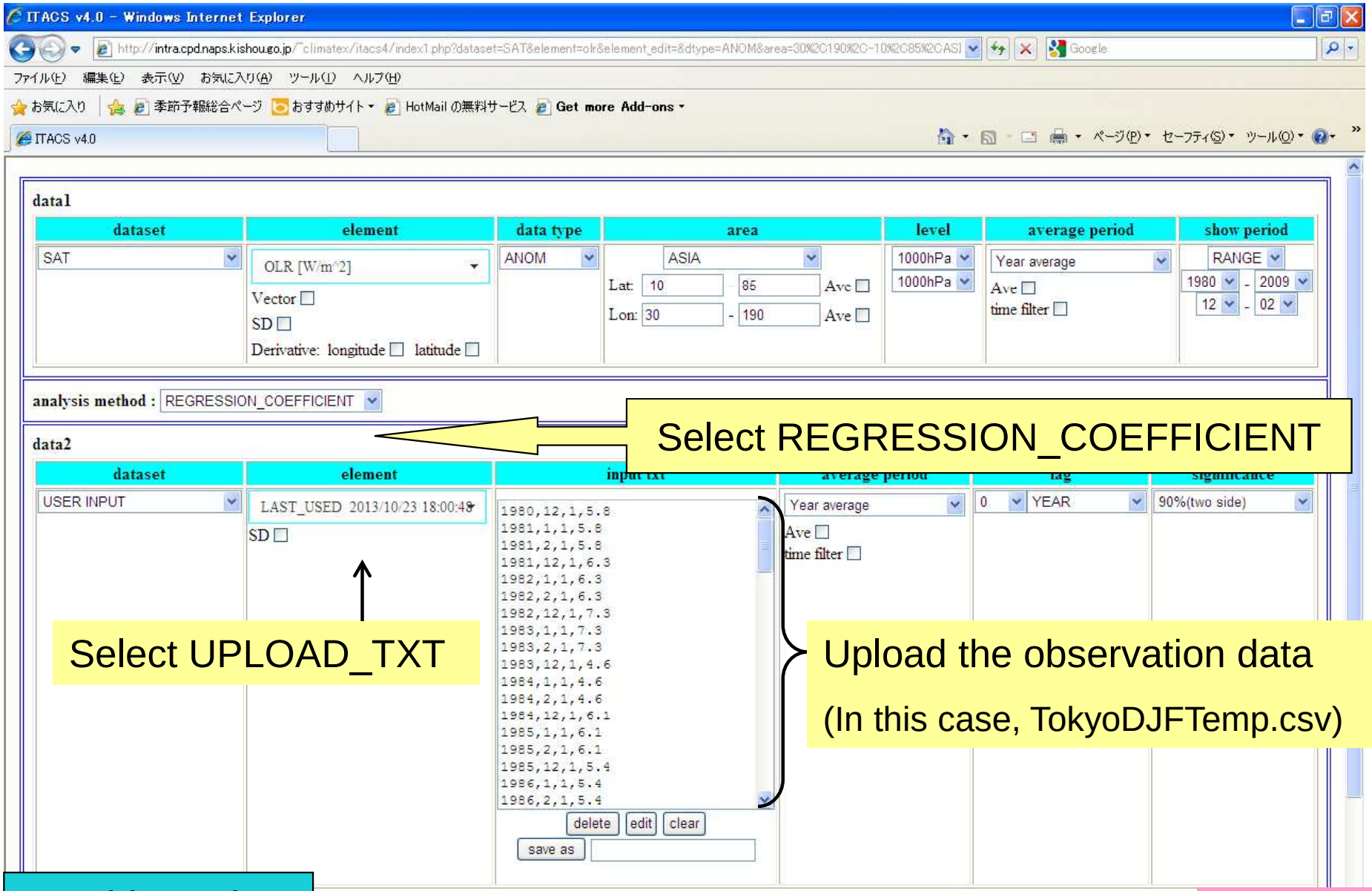
Temperature / Precipitation

Select the most effective combination of predictors after trial and error.

Correlation coefficients (the absolute values)

The ITACS will help you find them.

2. Multiple Regression Model

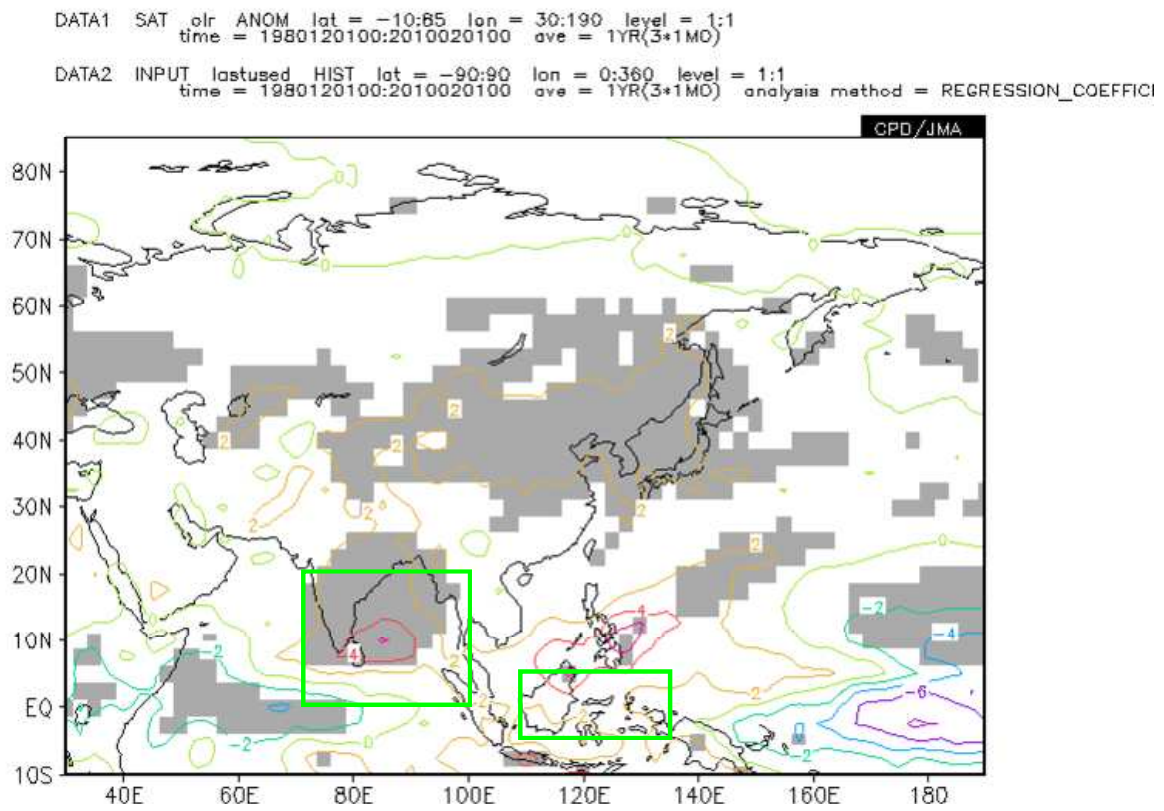


Graphic Option

Drawing: CONTOUR

ITACS

2. Multiple Regression Model



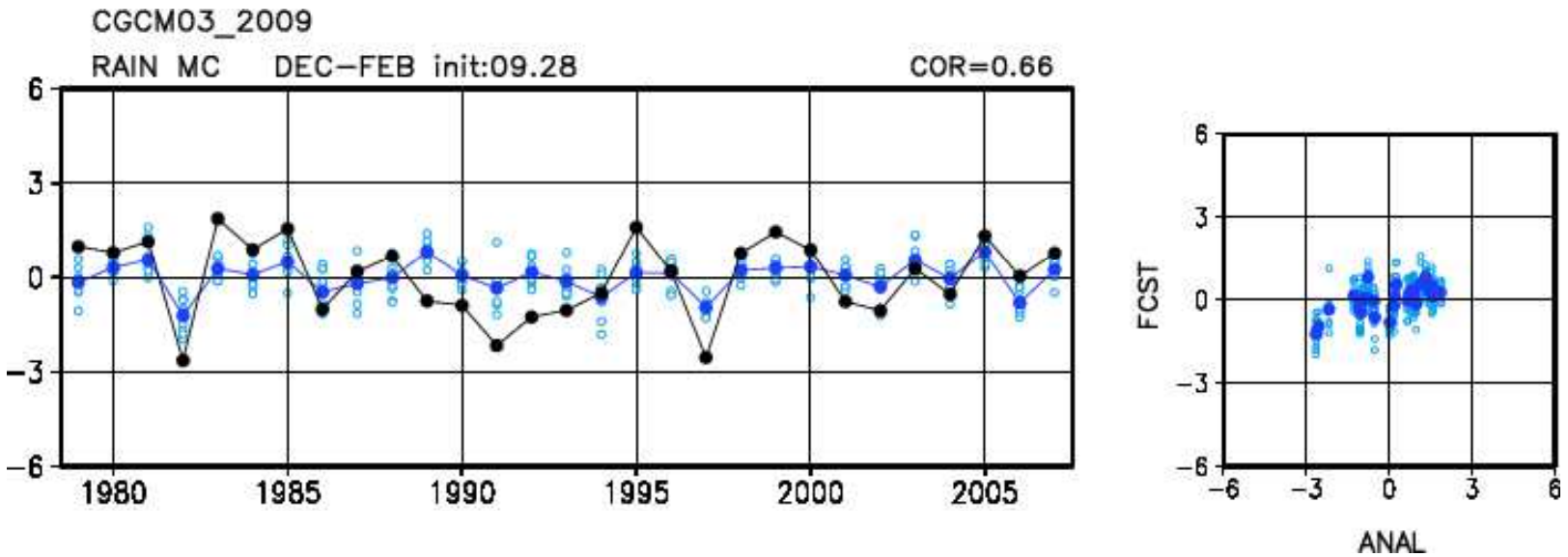
This is the relationship between OLR and DJF temperature in Tokyo. The OLR (convective activity) in Maritime Continent and that in the eastern Indian Ocean look possible predictors.

However, it is necessary to confirm the seasonal forecast model's skill about them.

ITACS

2. Multiple Regression Model

Prediction skill evaluated by 30 years hindcast

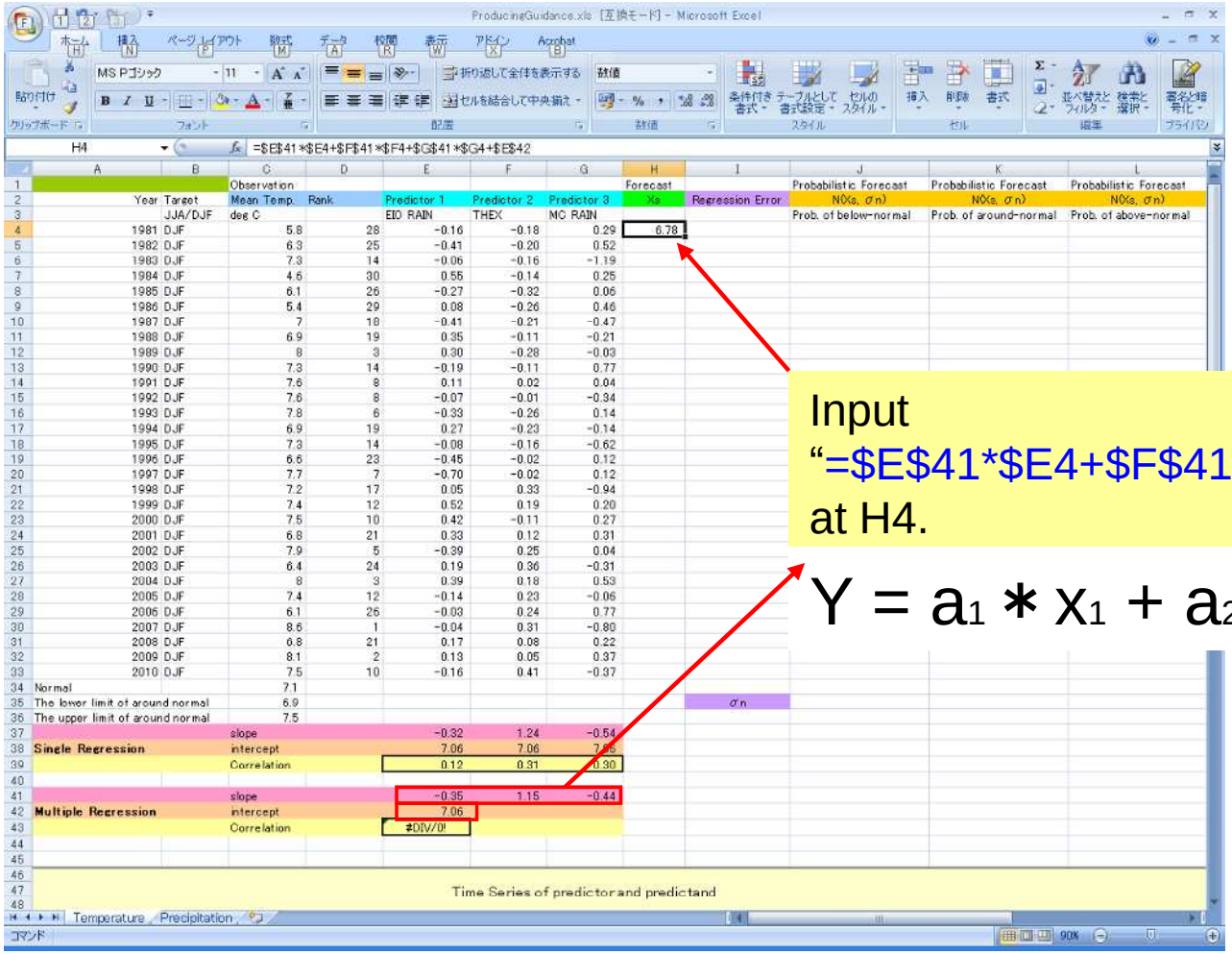


Anomaly correlation between the observed and forecasted DJF rainfall amounts in Maritime Continent is more than 0.6.
Prediction skill is good.

http://ds.data.jma.go.jp/tcc/tcc/products/model/hindcast/cgcm03_vrf/shisuu_10mem12mon_tcc.html

2. Multiple Regression Model

- Calculate the forecasts using multiple regression equation.

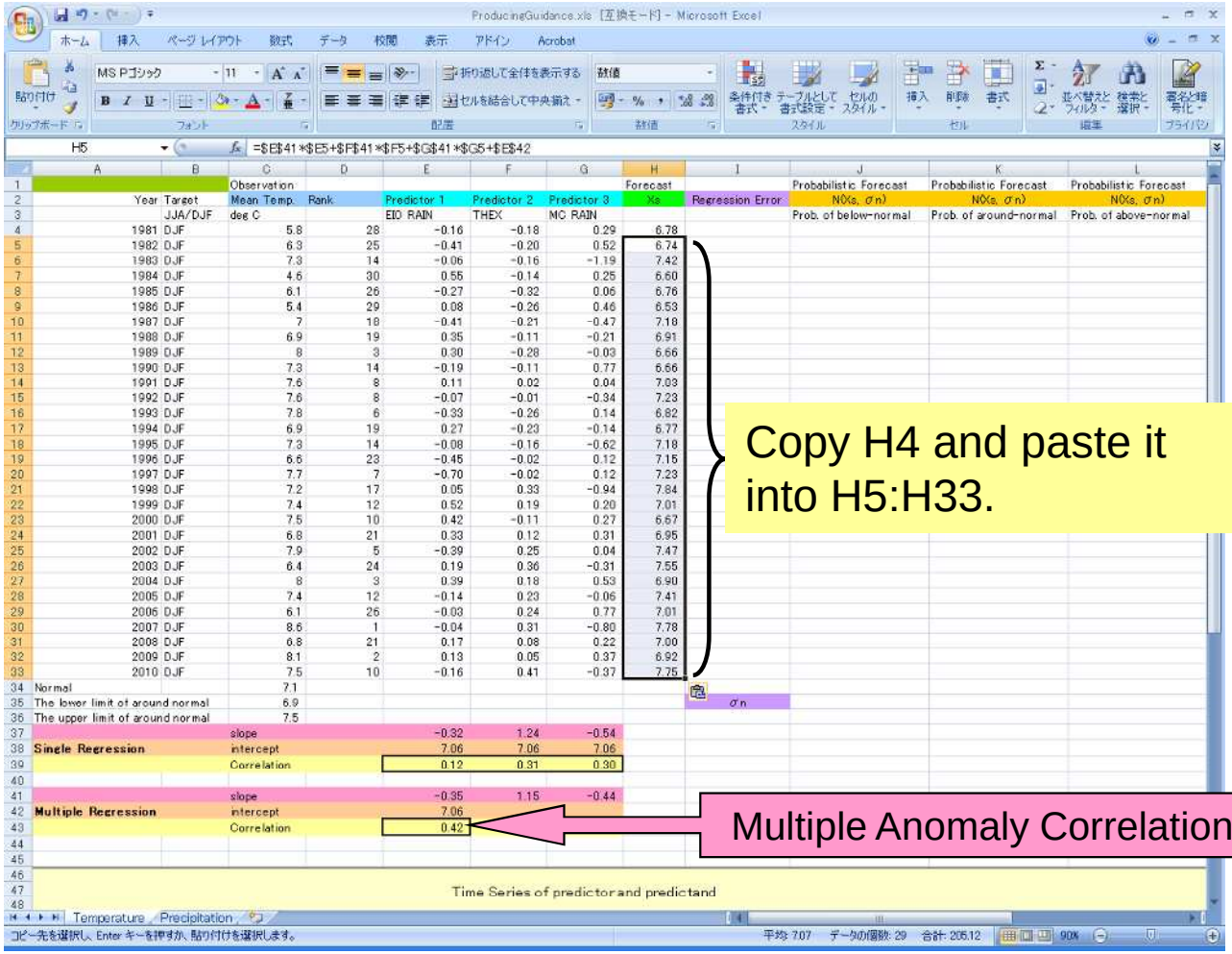


Input
"=$E$41*$E4+$F$41*$F4+$G$41*$G4+$E$42"
at H4.

$$Y = a_1 * x_1 + a_2 * x_2 + a_3 * x_3 + b$$

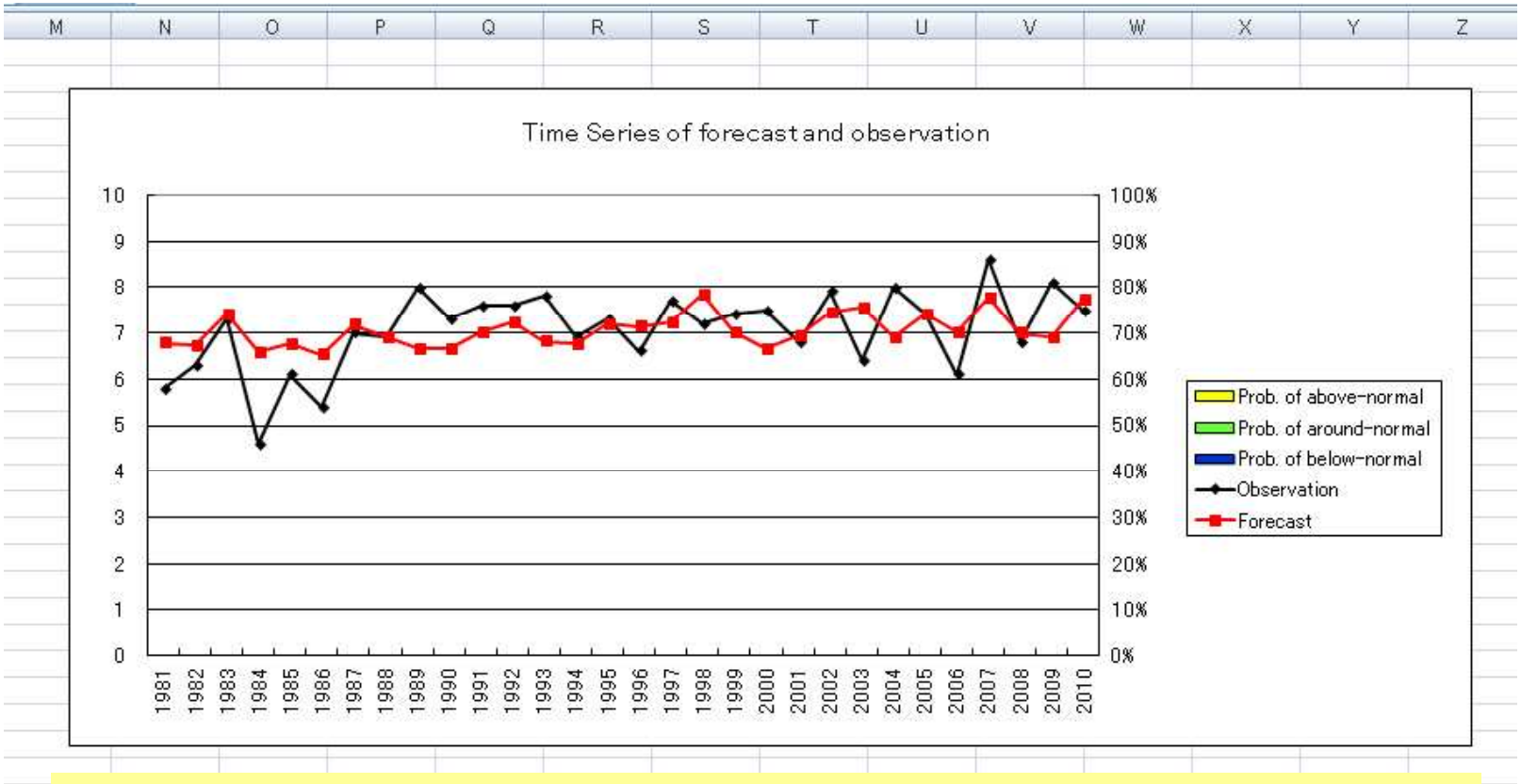
2. Multiple Regression Model

- Calculate the forecasts using multiple regression equation.



2. Multiple Regression Model

The time series of the forecast data has been automatically updated.



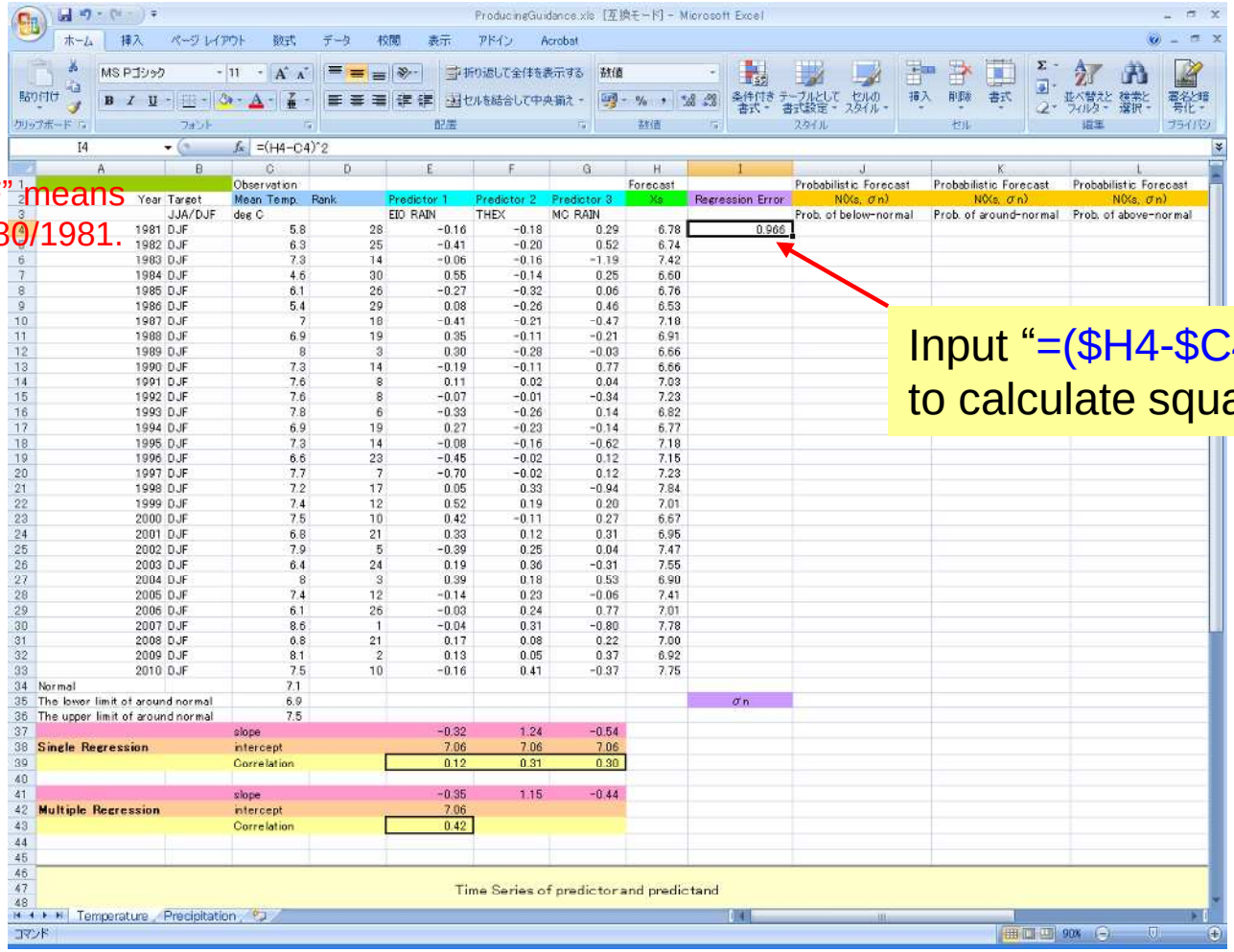
I will ask you the following questions at the end of this session:

1. “What predictors did you select?
2. “Is the multiple regression model guidance more accurate than the single regression model guidance?”

3. Probabilistic Forecast

- Calculate squares of regression errors.
- Calculate the root mean square error.

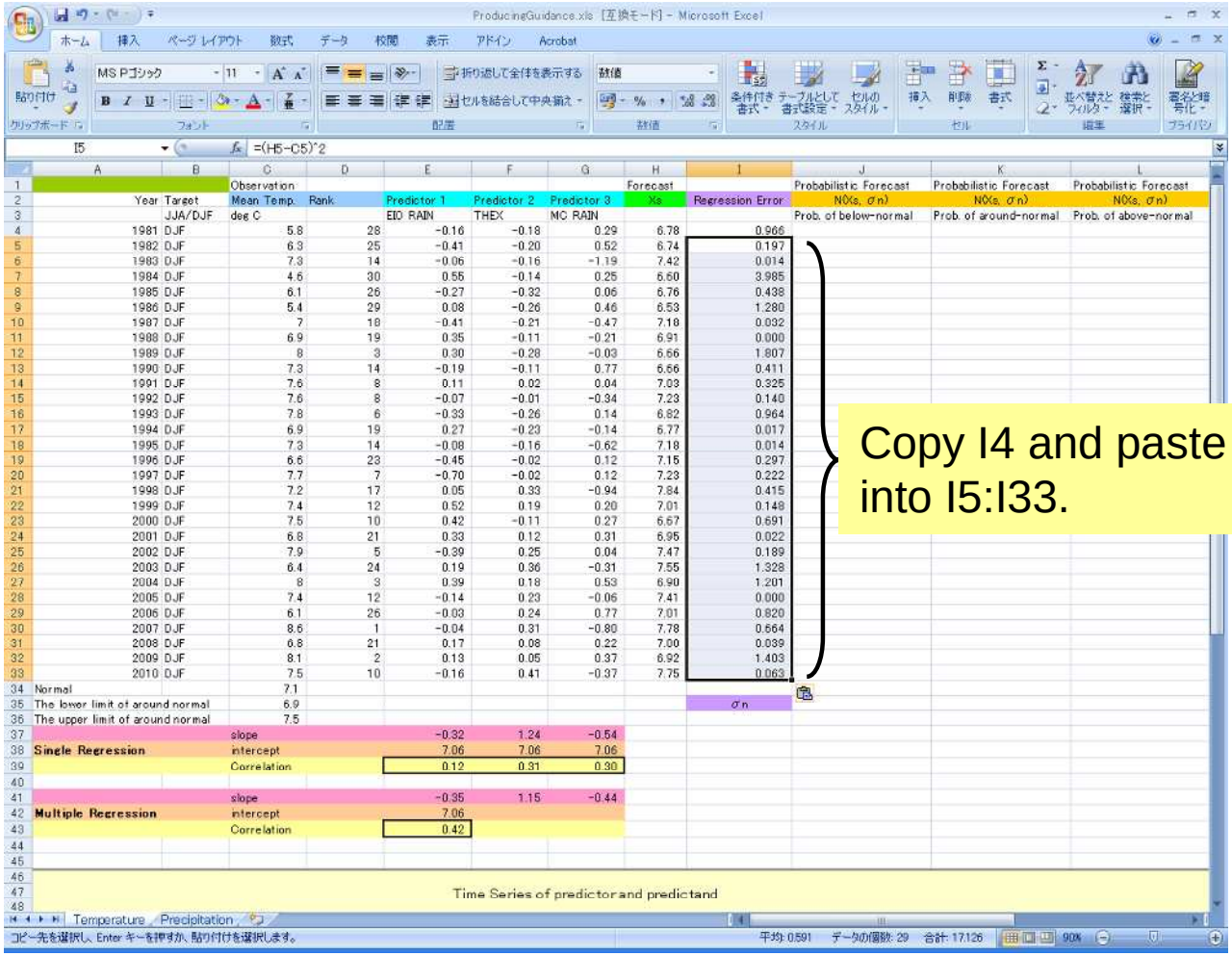
“1981 DJF” means DJF of 1980/1981.



Input “=($H4-$C4)^2” at I4 to calculate square error.

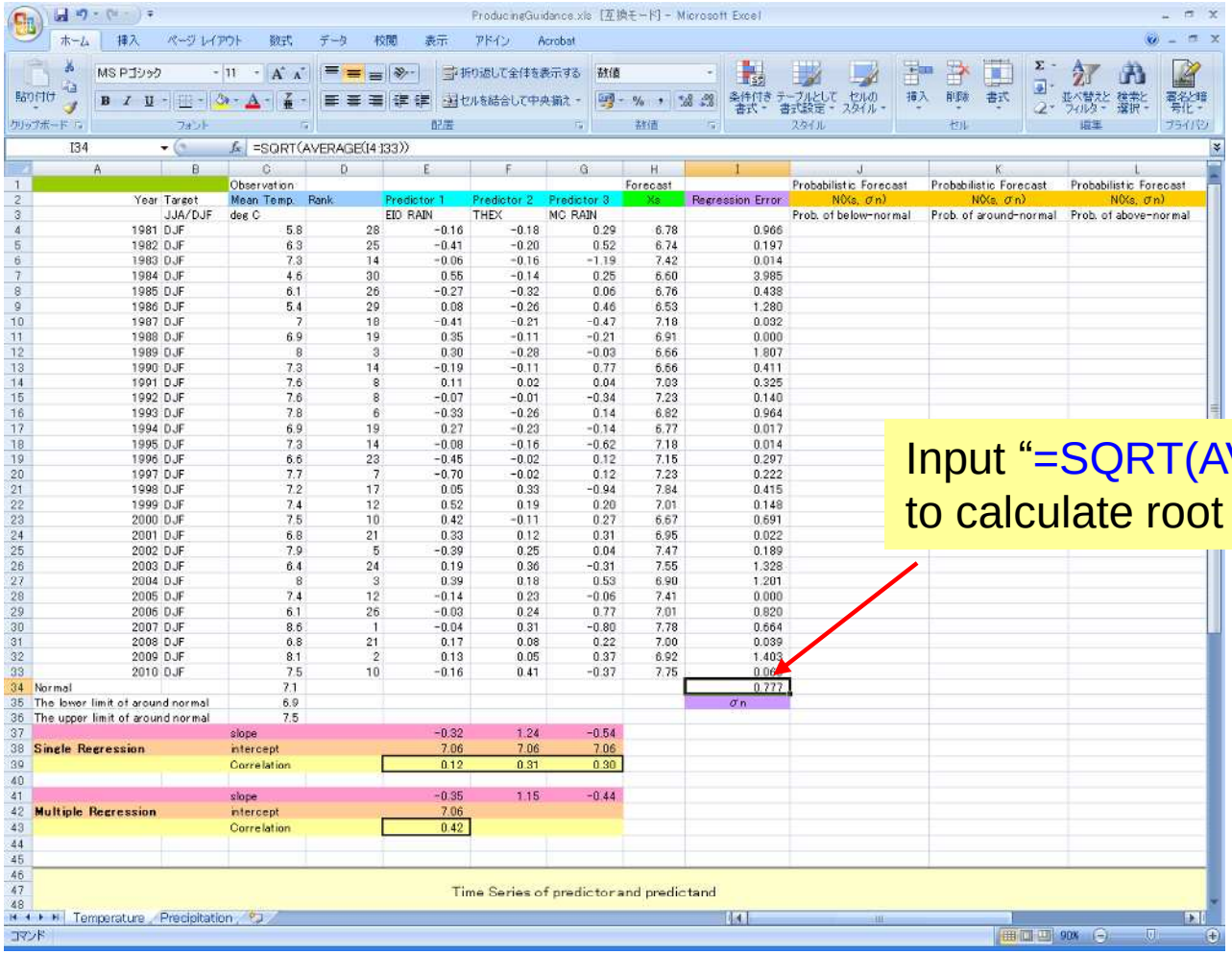
3. Probabilistic Forecast

- Calculate squares of regression errors.
- Calculate the root mean square error.



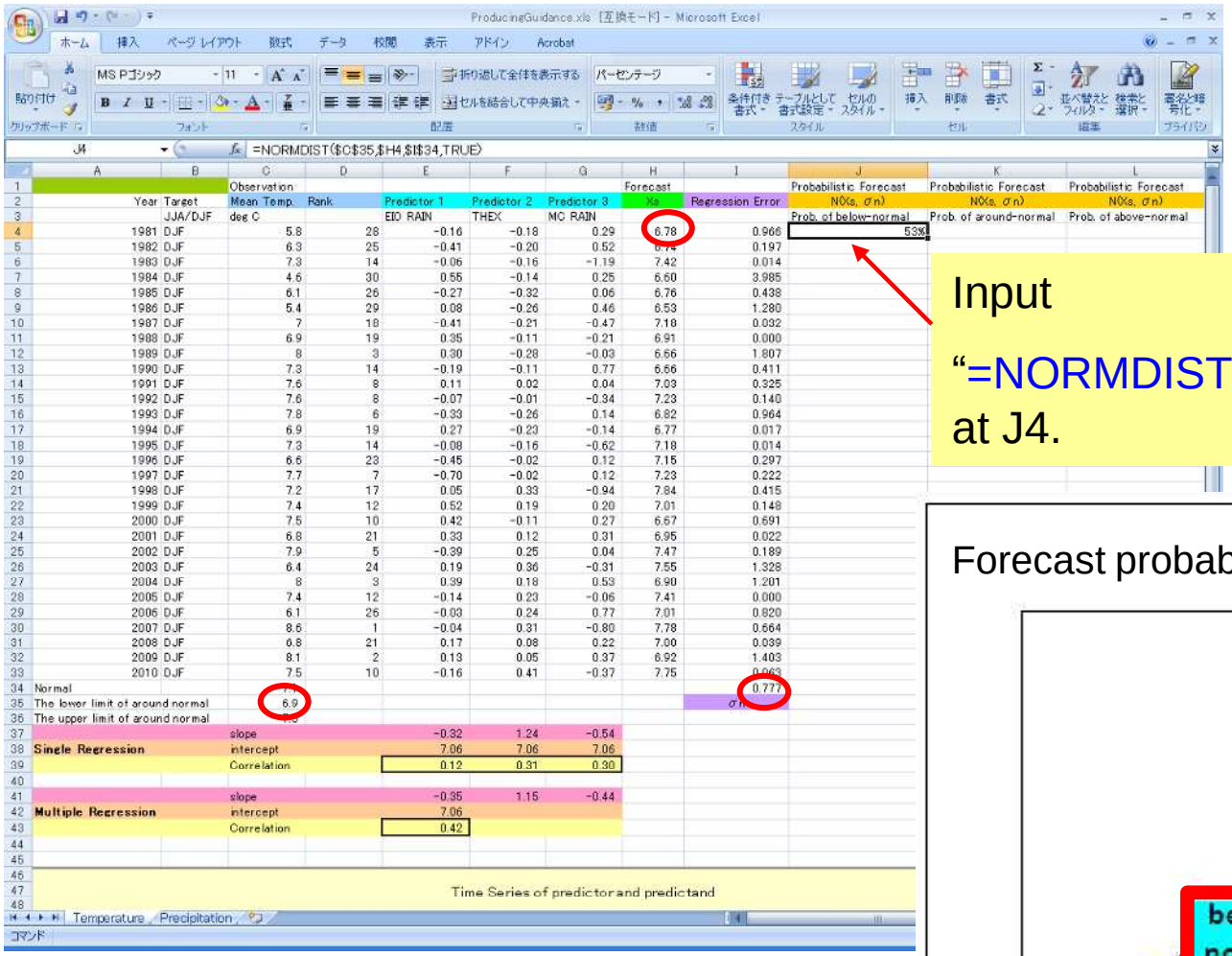
3. Probabilistic Forecast

- Calculate squares of regression errors.
- Calculate the root mean square error.



3. Probabilistic Forecast

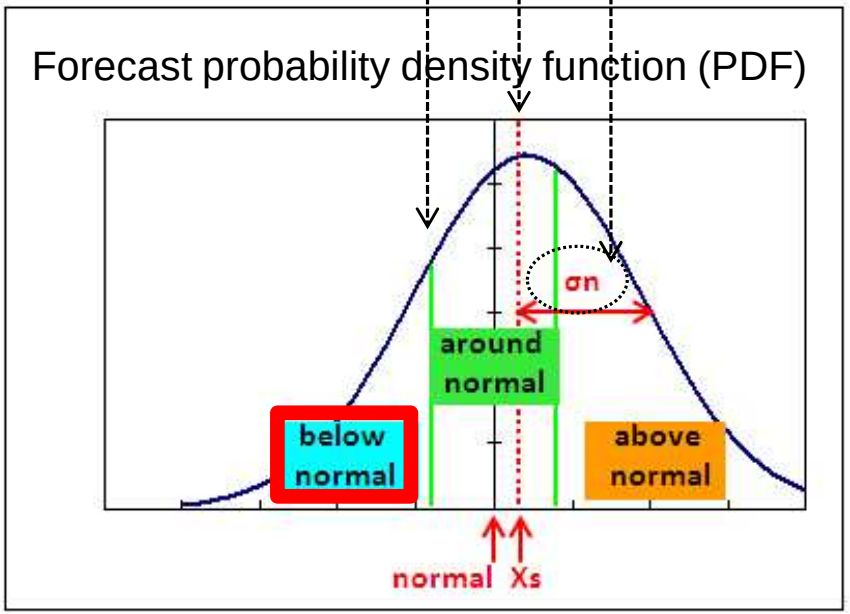
- Calculate probability of below-normal.



TRUE:
Value of cumulative distribution function for the first argument (i.e. $C$35).
FALSE:
Value of probability density function for the first argument (i.e. $C$35).

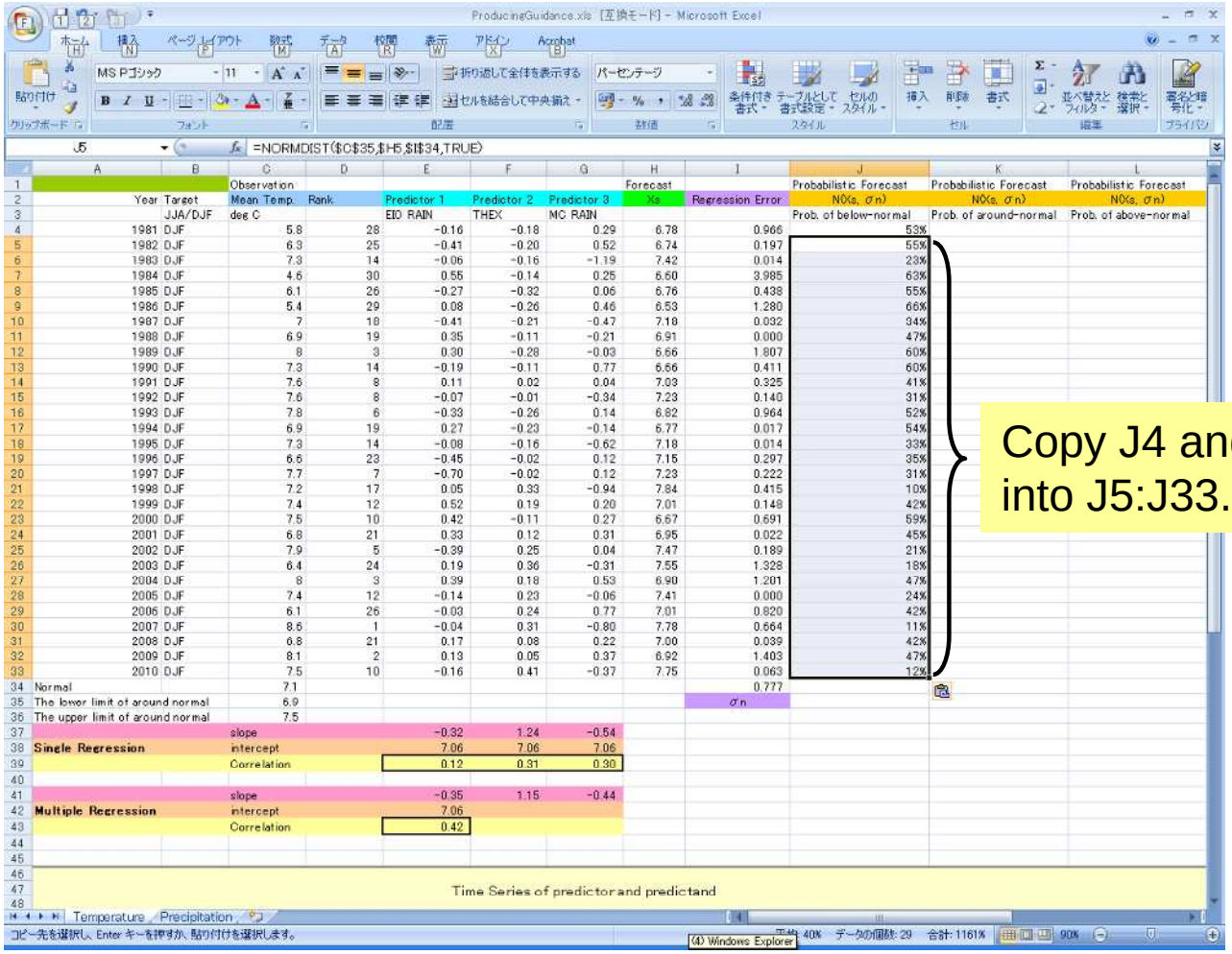
Input

"=NORMDIST($C$35,$H4,$I$34,TRUE)"
at J4.



3. Probabilistic Forecast

- Calculate probability of below-normal.



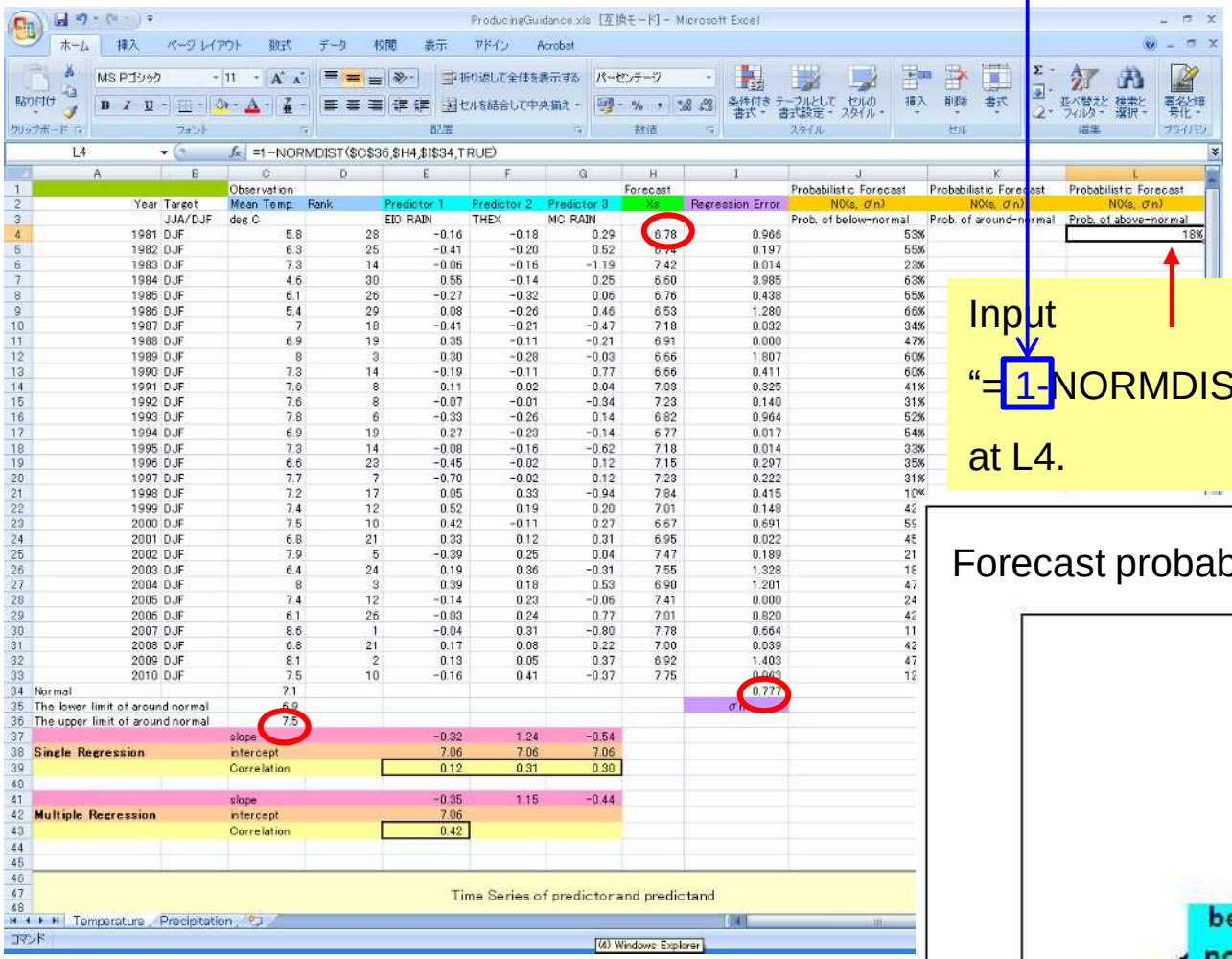
3. Probabilistic Forecast

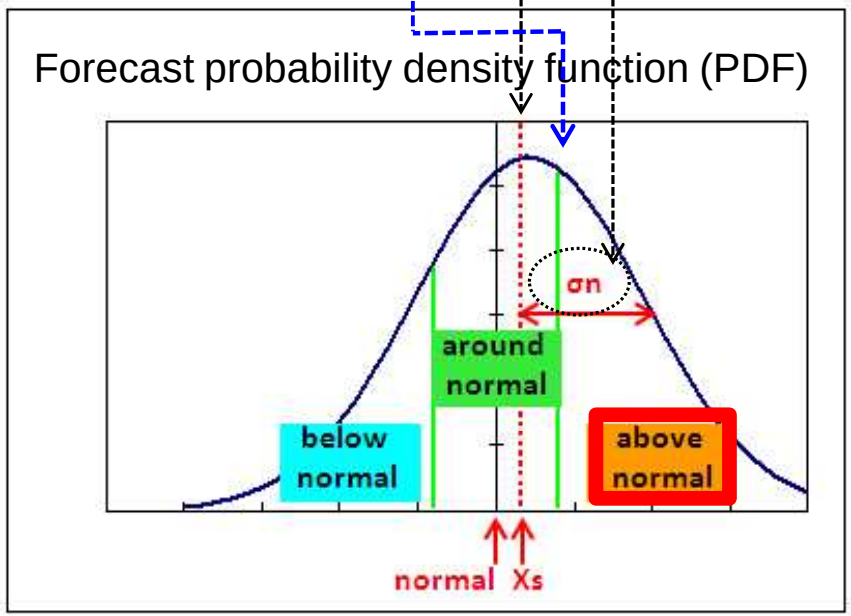
- Calculate probability of above-normal.

Very important!

TRUE:
Value of cumulative distribution function for the first argument (i.e. $C$36).
FALSE:
Value of probability density function for the first argument (i.e. $C$36).

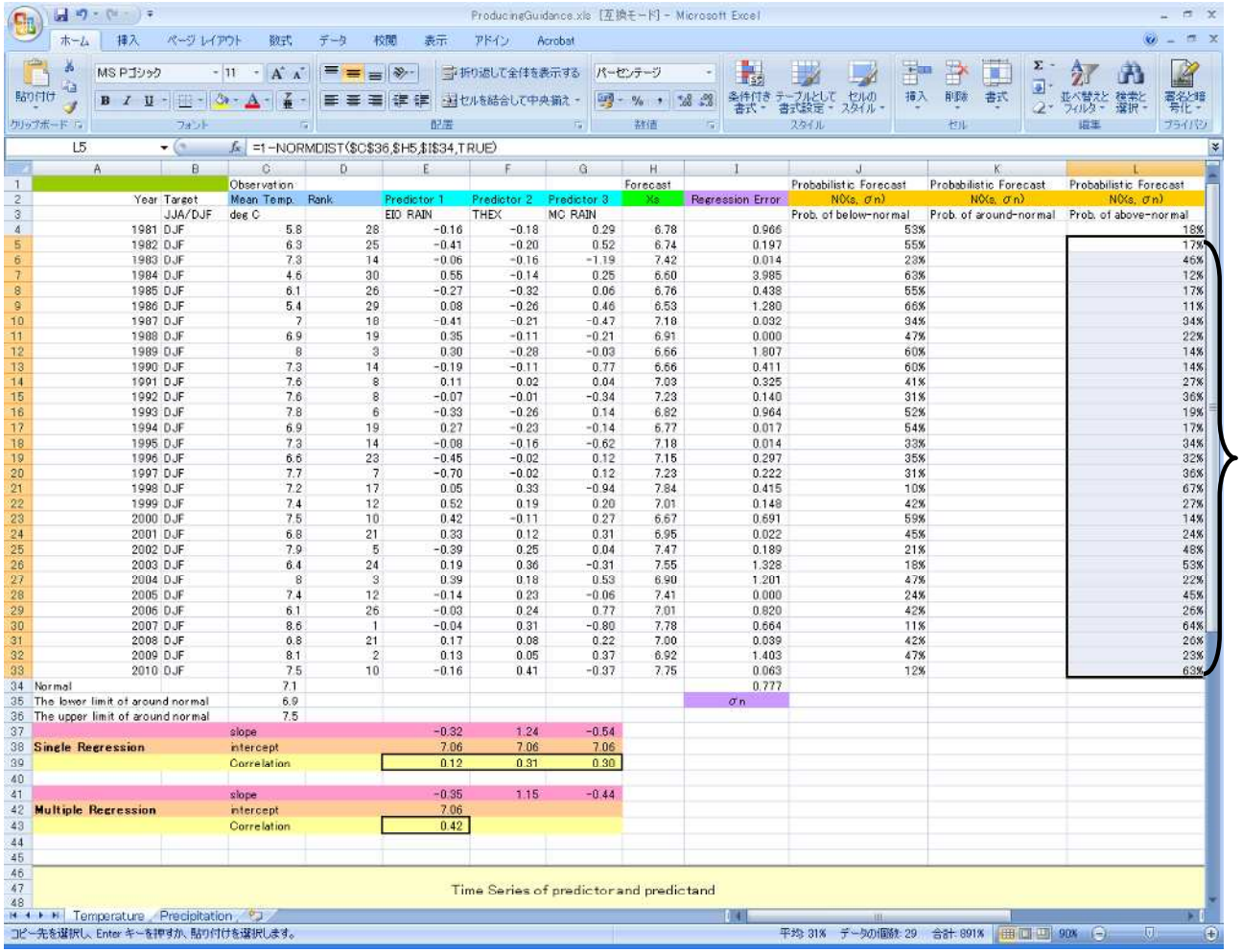
Input
"=1-NORMDIST($C$36,$H4,$I$34,TRUE)"
at L4.





3. Probabilistic Forecast

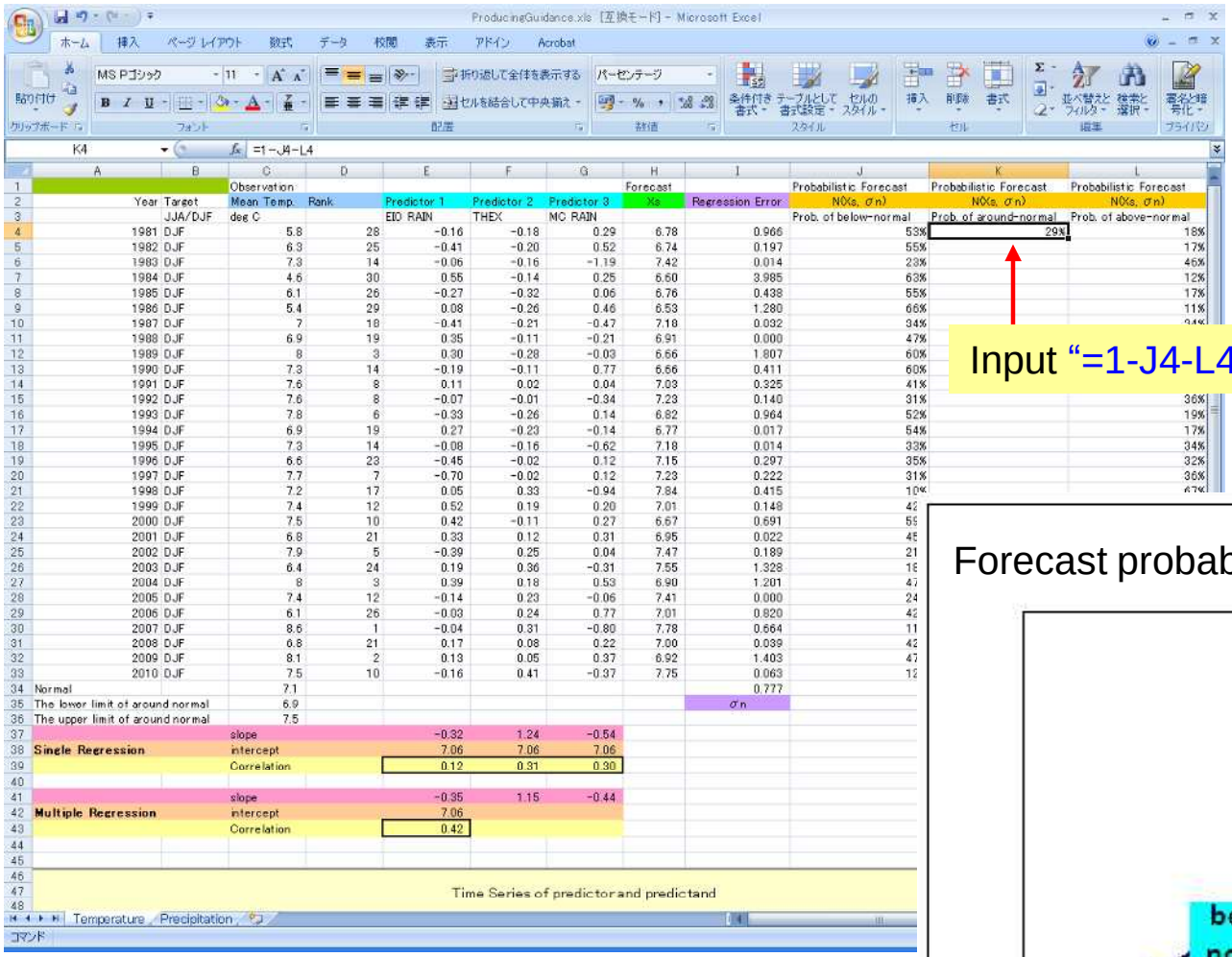
- Calculate probability of above-normal.



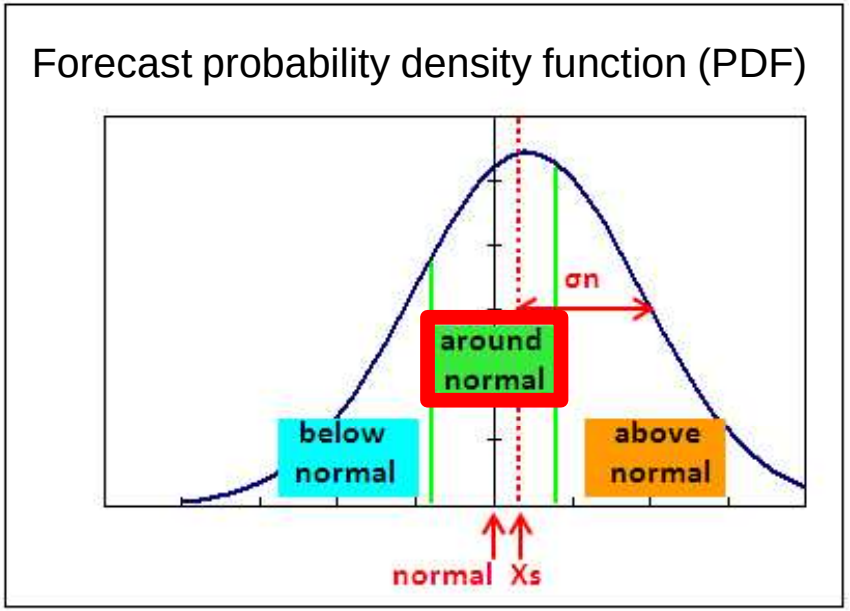
Copy L4 and paste it into L5:L33.

3. Probabilistic Forecast

- Calculate probability of around-normal.

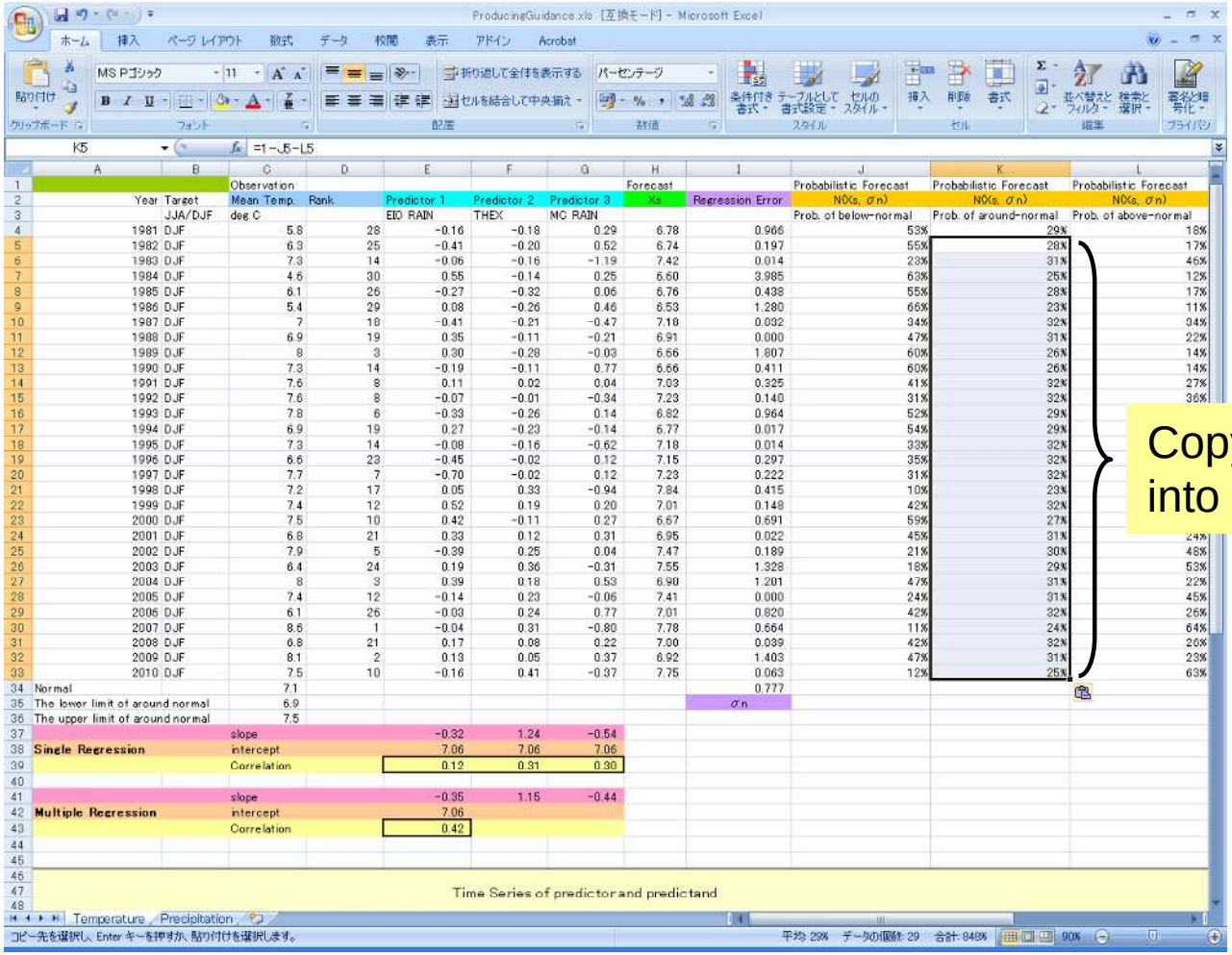


Input “=1-J4-L4“ at K4.



3. Probabilistic Forecast

- Calculate probability of around-normal.

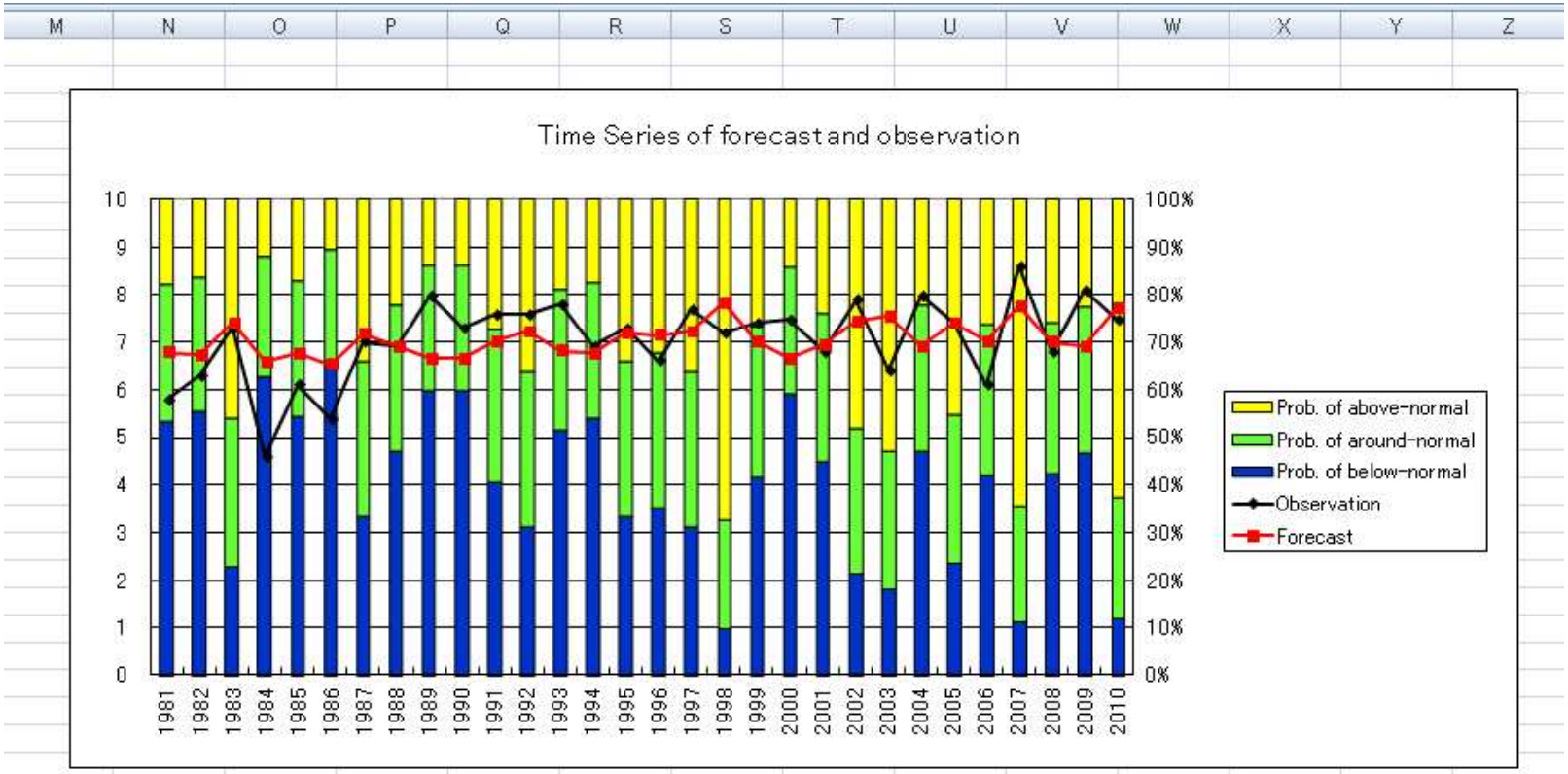


	A	B	C	D	E	F	G	H	I	J	K	L
1			Observation					Forecast		Probabilistic Forecast	Probabilistic Forecast	Probabilistic Forecast
2	Year	Target	Mean Temp.	Rank	Predictor 1	Predictor 2	Predictor 3	Xa	Regression Error	N(Xa, σn)	N(Xa, σn)	N(Xa, σn)
3		JJA/DJF	deg C		EIO RAIN	THEX	MC RAIN			Prob. of below-normal	Prob. of around-normal	Prob. of above-normal
4	1981	DJF	5.8	28	-0.16	-0.18	0.29	6.78	0.966	53%	29%	18%
5	1982	DJF	6.3	25	-0.41	-0.20	0.52	6.74	0.197	55%	28%	17%
6	1983	DJF	7.3	14	-0.06	-0.16	-1.19	7.42	0.014	23%	31%	46%
7	1984	DJF	4.6	30	0.55	-0.14	0.25	6.60	3.985	63%	25%	12%
8	1985	DJF	6.1	26	-0.27	-0.32	0.06	6.76	0.438	55%	28%	17%
9	1986	DJF	5.4	29	0.08	-0.26	0.46	6.53	1.280	66%	23%	11%
10	1987	DJF	7	18	-0.41	-0.21	-0.47	7.18	0.032	34%	32%	34%
11	1988	DJF	6.9	19	0.35	-0.11	-0.21	6.91	0.000	47%	31%	22%
12	1989	DJF	8	3	0.30	-0.28	-0.03	6.66	1.807	60%	26%	14%
13	1990	DJF	7.3	14	-0.19	-0.11	0.77	6.66	0.411	60%	26%	14%
14	1991	DJF	7.6	8	0.11	0.02	0.04	7.03	0.325	41%	32%	27%
15	1992	DJF	7.6	8	-0.07	-0.01	-0.34	7.23	0.140	31%	32%	36%
16	1993	DJF	7.8	6	-0.33	-0.26	0.14	6.82	0.964	52%	29%	
17	1994	DJF	6.9	19	0.27	-0.23	-0.14	6.77	0.017	54%	29%	
18	1995	DJF	7.3	14	-0.08	-0.16	-0.62	7.18	0.014	33%	32%	
19	1996	DJF	6.6	23	-0.45	-0.02	0.12	7.15	0.297	35%	32%	
20	1997	DJF	7.7	7	-0.70	-0.02	0.12	7.23	0.222	31%	32%	
21	1998	DJF	7.2	17	0.05	0.33	-0.94	7.84	0.415	10%	23%	
22	1999	DJF	7.4	12	0.52	0.19	0.20	7.01	0.148	42%	32%	
23	2000	DJF	7.5	10	0.42	-0.11	0.27	6.57	0.691	59%	27%	
24	2001	DJF	6.8	21	0.33	0.12	0.31	6.95	0.022	45%	31%	24%
25	2002	DJF	7.9	5	-0.39	0.25	0.04	7.47	0.189	21%	30%	48%
26	2003	DJF	6.4	24	0.19	0.36	-0.31	7.55	1.328	18%	29%	53%
27	2004	DJF	8	3	0.39	0.18	0.53	6.90	1.201	47%	31%	22%
28	2005	DJF	7.4	12	-0.14	0.23	-0.06	7.41	0.000	24%	31%	45%
29	2006	DJF	6.1	26	-0.03	0.24	0.77	7.01	0.820	42%	32%	26%
30	2007	DJF	8.6	1	-0.04	0.31	-0.80	7.78	0.664	11%	24%	64%
31	2008	DJF	6.8	21	0.17	0.08	0.22	7.00	0.039	42%	32%	26%
32	2009	DJF	8.1	2	0.13	0.05	0.37	6.92	1.403	47%	31%	23%
33	2010	DJF	7.5	10	-0.16	0.41	-0.37	7.75	0.063	12%	25%	63%
34	Normal		7.1						0.777			
35	The lower limit of around normal		6.9						σn			
36	The upper limit of around normal		7.5									
37			slope		-0.32	1.24	-0.54					
38	Single Regression		intercept		7.06	7.06	7.06					
39			Correlation		0.12	0.31	0.30					
40												
41			slope		-0.35	1.15	-0.44					
42	Multiple Regression		intercept		7.06							
43			Correlation		0.42							
44												
45												
46												
47					Time Series of predictor and predictand							
48												

Copy K4 and paste it into K5:K33.

3. Probabilistic Forecast

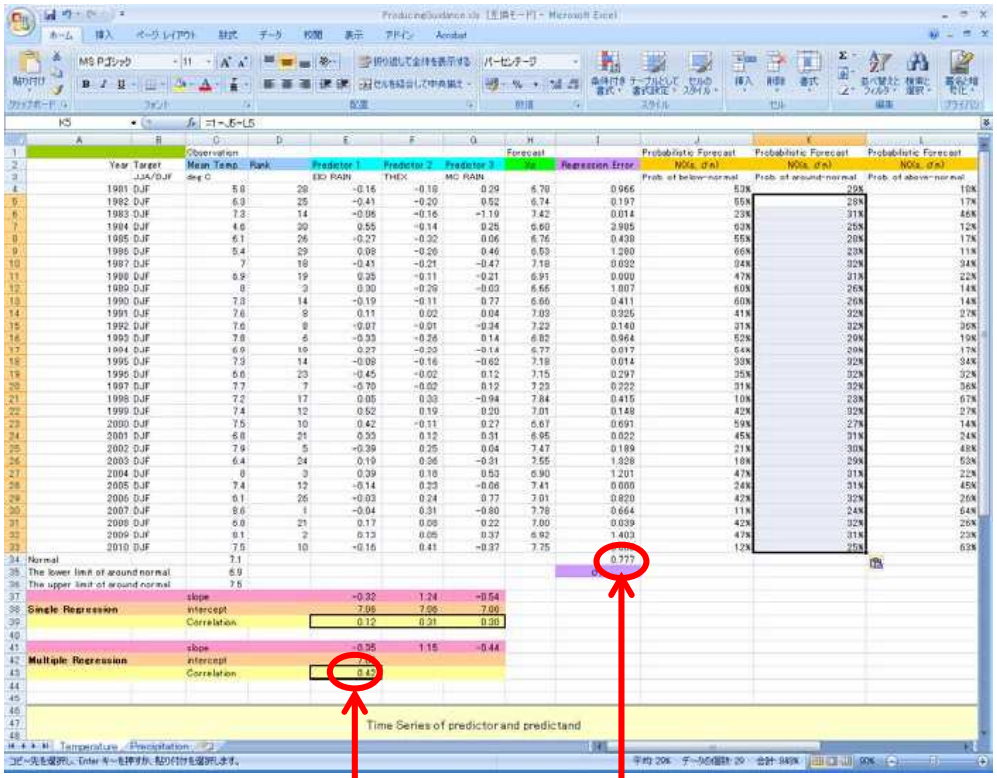
The time series of the probabilistic forecast data has been automatically added.

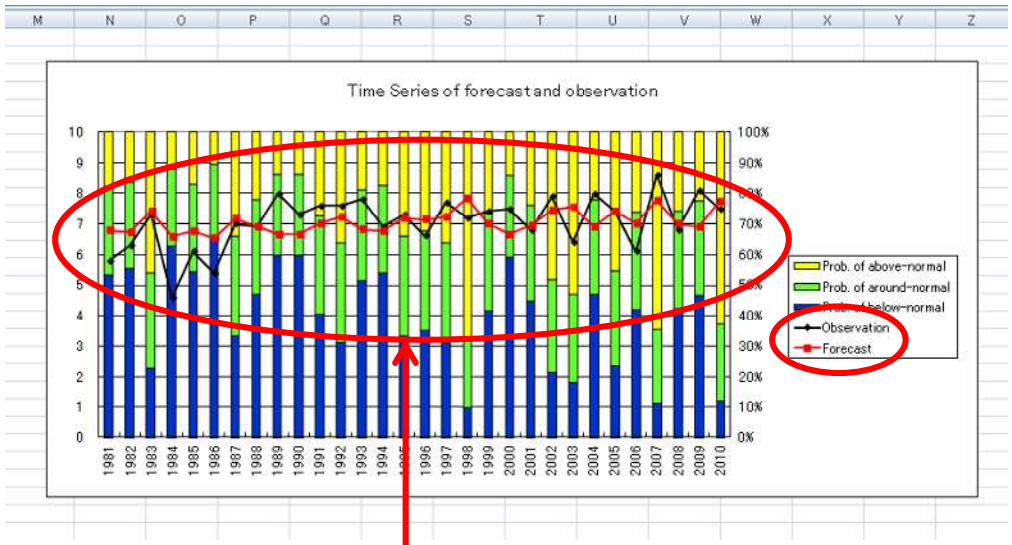


I will ask you the following questions at the end of this session:

1. “How did the guidance predict the hottest/coldest/wettest/driest year in your country?”
2. “Are there any essential differences of the probability between the temperature and precipitation?”

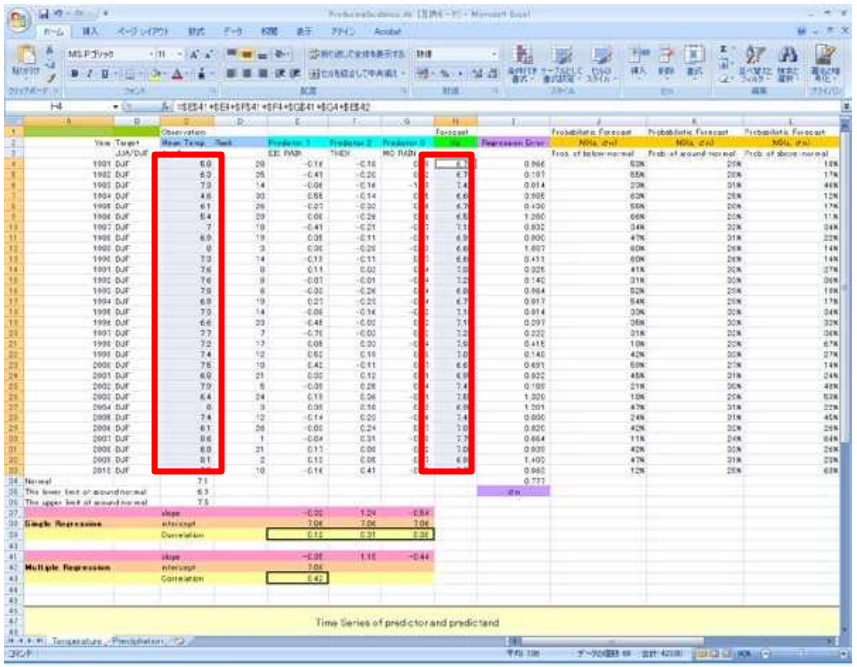
4-1. Deterministic Verification

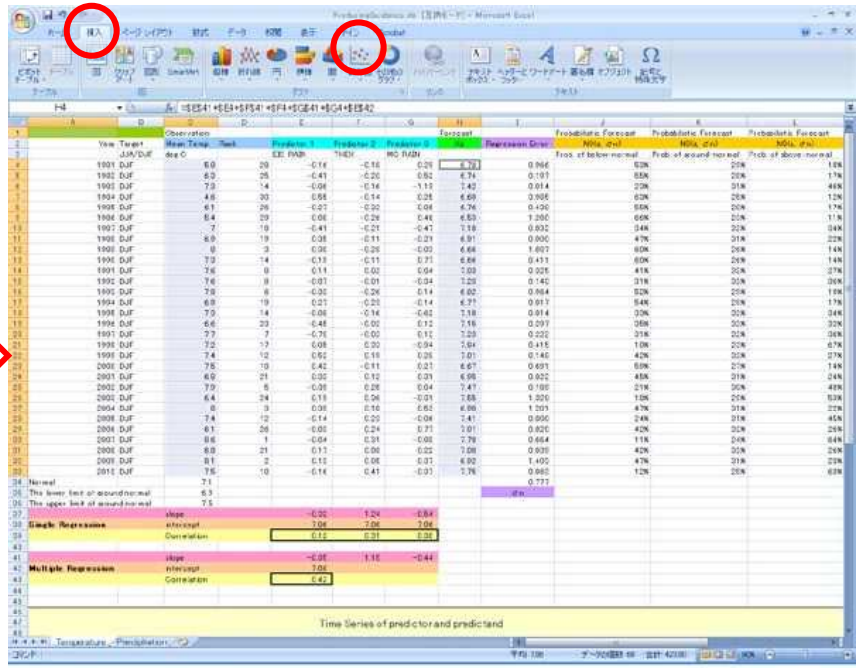




- **Anomaly Correlation**
- **Root Mean Square Error (RMSE)**
- **Time Series Chart**
- **Scatter Plot**

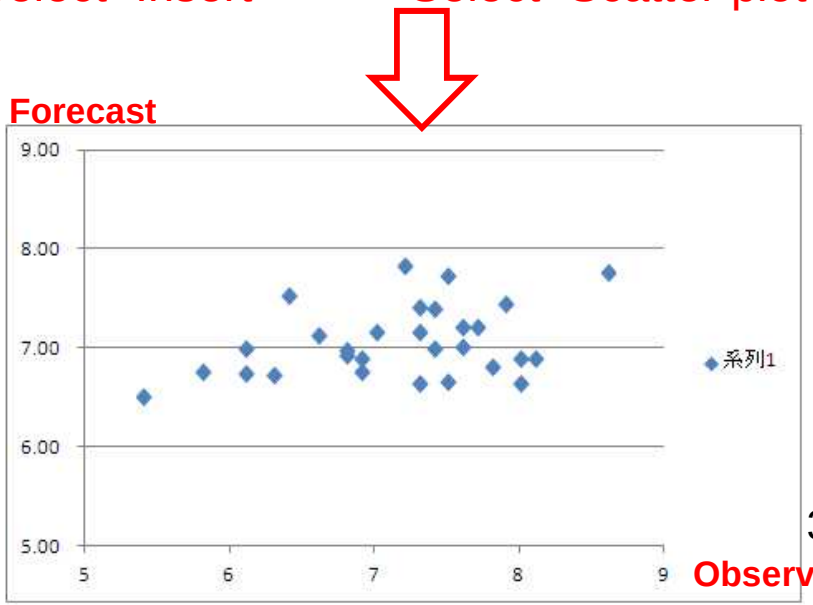
4-1. Deterministic Verification





Select C4:C33 ⟶ Select H4:H33
(with pushing “ctrl” key)

Select “Insert” ⟶ Select “Scatter plot”



- Anomaly Correlation
- Root Mean Square Error (RMSE)
- Time Series Chart
- **Scatter Plot**

4-2. Probabilistic Verification

Brier skill score (3-category forecast)

- Brier score measures mean squared error of the probability forecasts.

$$BS = \frac{1}{2N}\sum_{i=1}^{N}\sum_{m=1}^{3}(p_i^m - o_i^m)^2$$

p_i^m : forecast probability

o_i^m : observed occurrence (0 or 1)

N : forecast frequency

m : category

Range: 0 to 1. Perfect score: 0

- Brier skill score measures ***skill*** relative to a reference forecast (usually climatology).

$$BSS = 1 - \frac{BS}{BS_{reference}}$$

Range: minus infinity to 1. BSS=0 indicates no skill when compared to the reference forecast. Perfect score: 1.

4-2. Probabilistic Verification

Brier skill score (3-category forecast)

- Brier skill score

  = the relative skill of the probabilistic forecast to the climatology

$$BSS = \frac{BS - BS_{c\lim}}{0 - BS_{c\lim}} = 1 - \frac{BS}{BS_{c\lim}}$$

$$Skill\ score = \frac{score_{forecast} - score_{reference}}{score_{perfect\ forecast} - score_{reference}}$$

$$BS_{c\lim} = 1/3$$

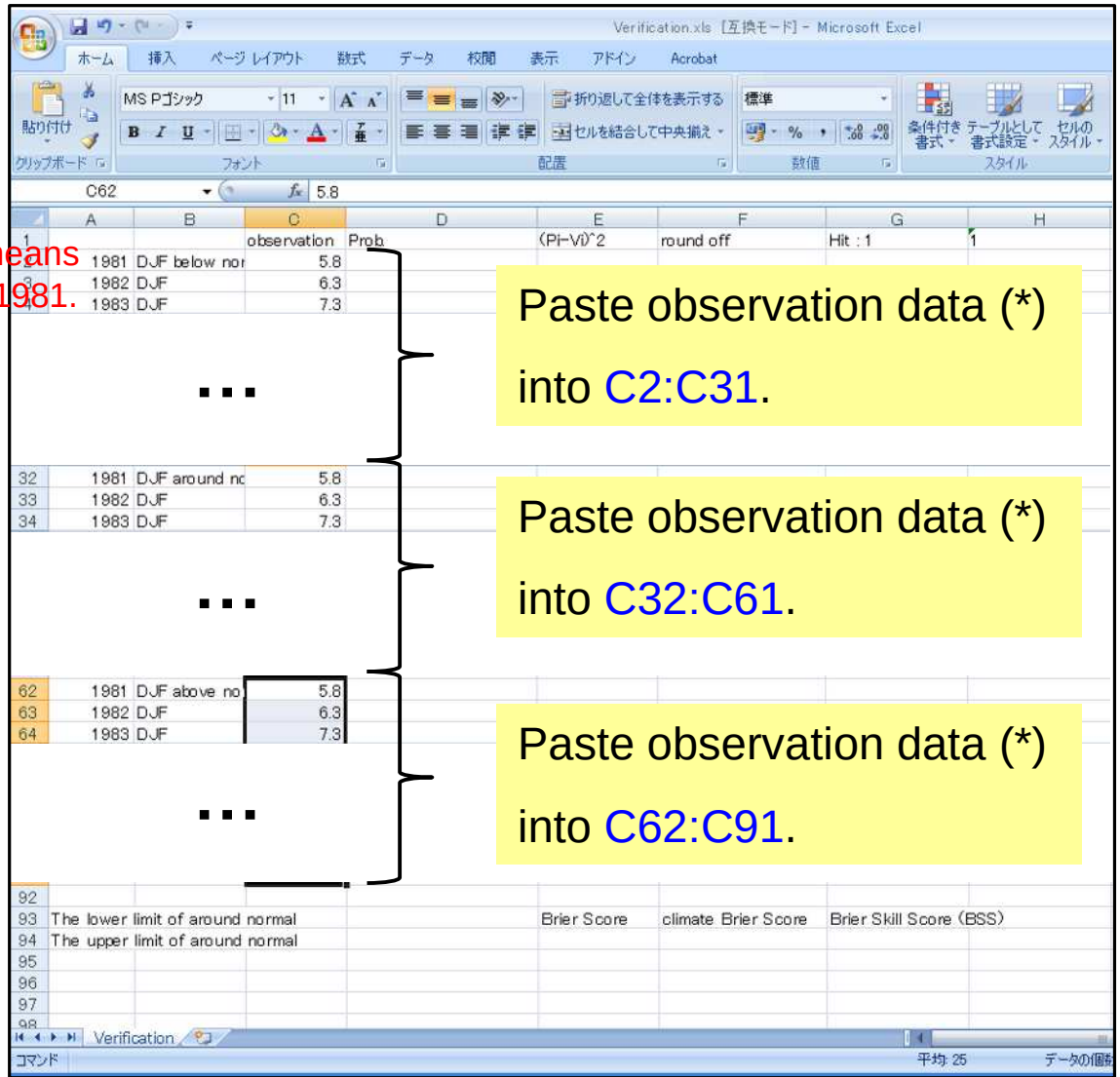
Perfect score: 1
BSS>0 : better than the climatology.
BSS=0 indicates no skill when compared to the climatology.
BSS<0 : worse than the climatology.

4-2. Probabilistic Verification

Brier skill score (3-category forecast)

•Open the Verification.xls.

＊ If you are busy, please use BriefVerification.xls in place of Verification.xls.



"1981 DJF" means DJF of 1980/1981.

Paste observation data (*) into C2:C31.

Paste observation data (*) into C32:C61.

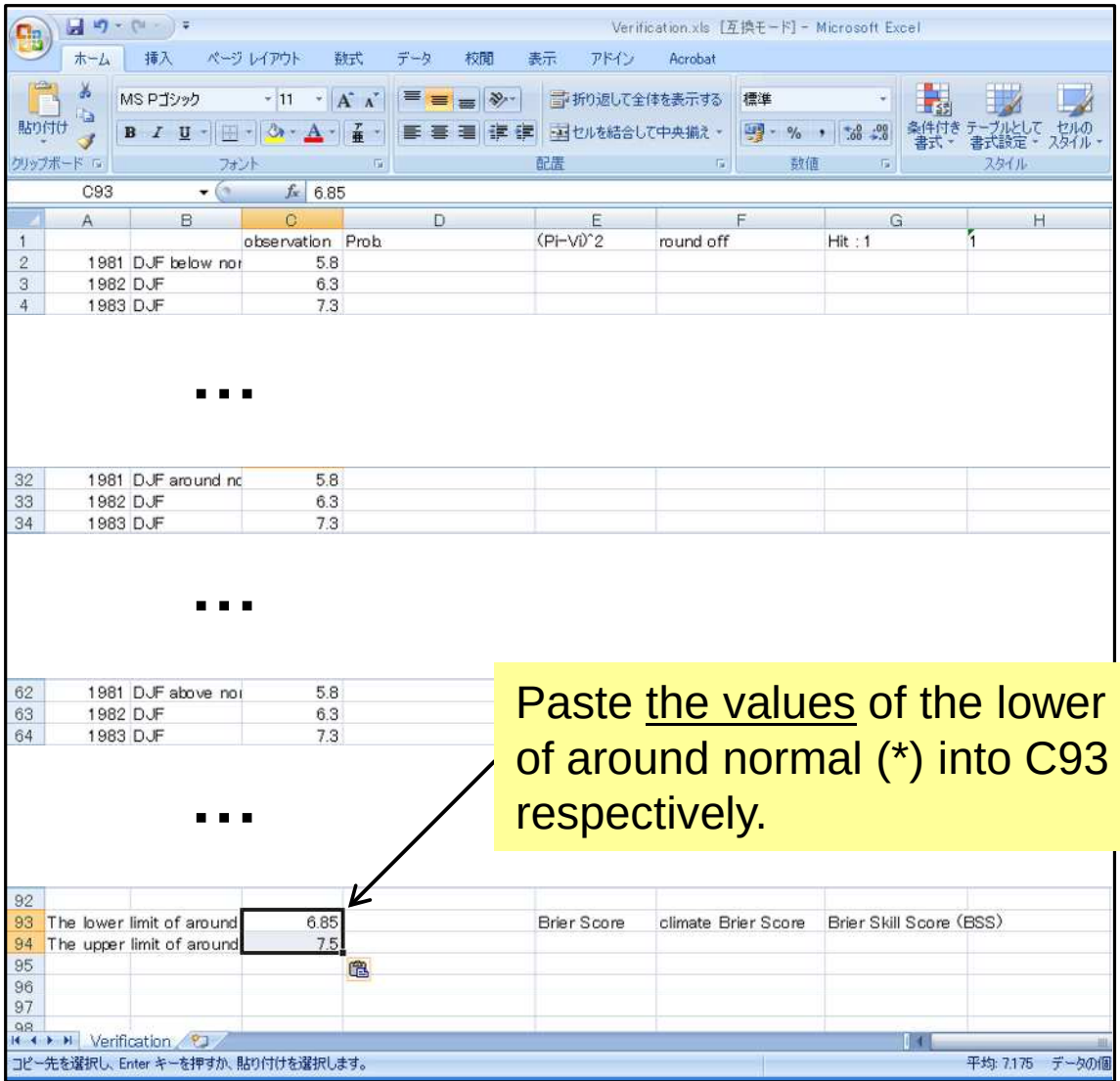
Paste observation data (*) into C62:C91.

In this case, the predictand is temperature and the station is Tokyo.

* These values should be copied from C4:C33 in the temperature sheet of the ProducingGuidance.xls.

4-2. Probabilistic Verification

Brier skill score (3-category forecast)

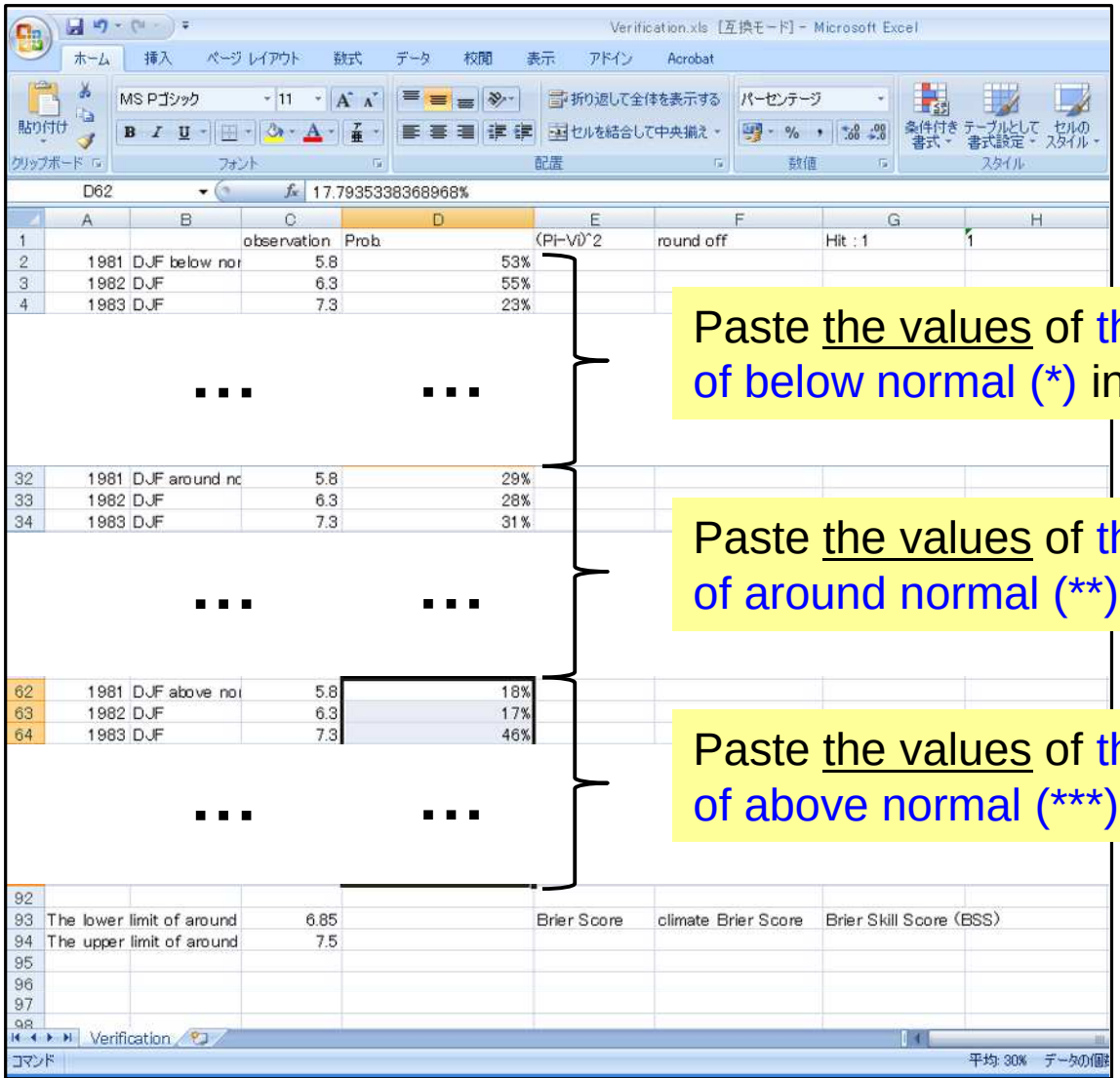


Paste the values of the lower and upper limits of around normal (*) into C93 and C94, respectively.

* These values should be copied from the C35 and C36 cells in the temperature sheet of the ProducingGuidance.xls

4-2. Probabilistic Verification

Brier skill score (3-category forecast)



Paste the values of the probabilities of below normal (*) into D2:D31.

* J4:J33 in the temperature sheet of the ProducingGuidance.xls

Paste the values of the probabilities of around normal (**) into D32:D61.

** K4:K33 in the temperature sheet of the ProducingGuidance.xls

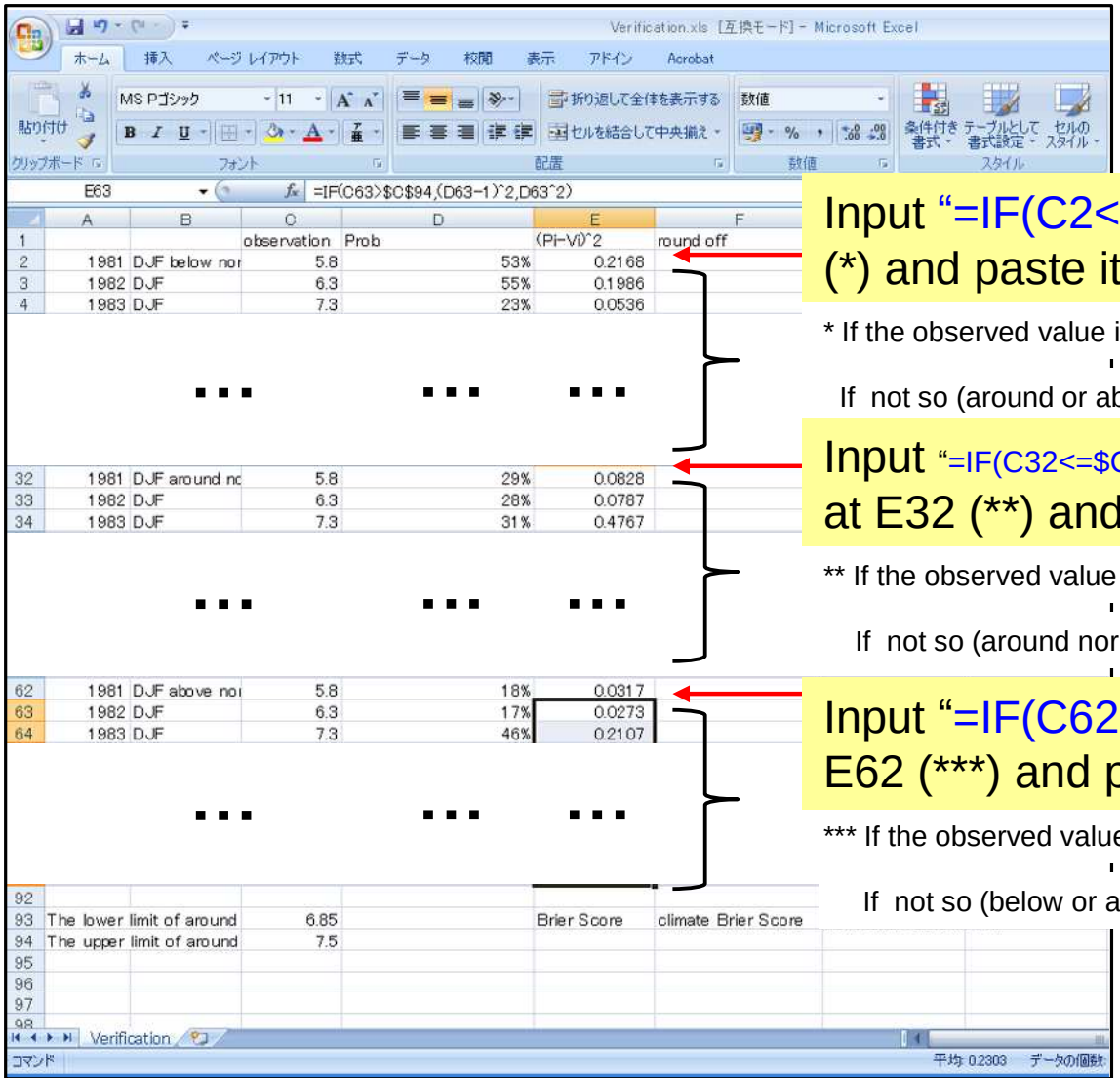
Paste the values of the probabilities of above normal (***) into D62:D91.

*** L4:L33 in the temperature sheet of the ProducingGuidance.xls

4-2. Probabilistic Verification

Brier skill score (3-category forecast)

* If you use BriefVerification.xls in place of Verification.xls, the procedure in this slide is unnecessary.



Input "=IF(C2<=$C$93,(D2-1)^2,D2^2)" at E2 (*) and paste it into E3:E31.

* If the observed value is below normal, $E2 = (p_i{}^m - 1)^2$

If not so (around or above normal), $E2 = (p_i{}^m - 0)^2$

Input "=IF(C32<=$C$93,D32^2,IF(C32>$C$94,D32^2, (D32-1)^2))" at E32 (**) and paste it into E33:E61.

** If the observed value is below or above normal, $E32 = (p_i{}^m - 0)^2$

If not so (around normal), $E32 = (p_i{}^m - 1)^2$

Input "=IF(C62>$C$94,(D62-1)^2,D62^2)" at E62 (***) and paste it into E63:E91.

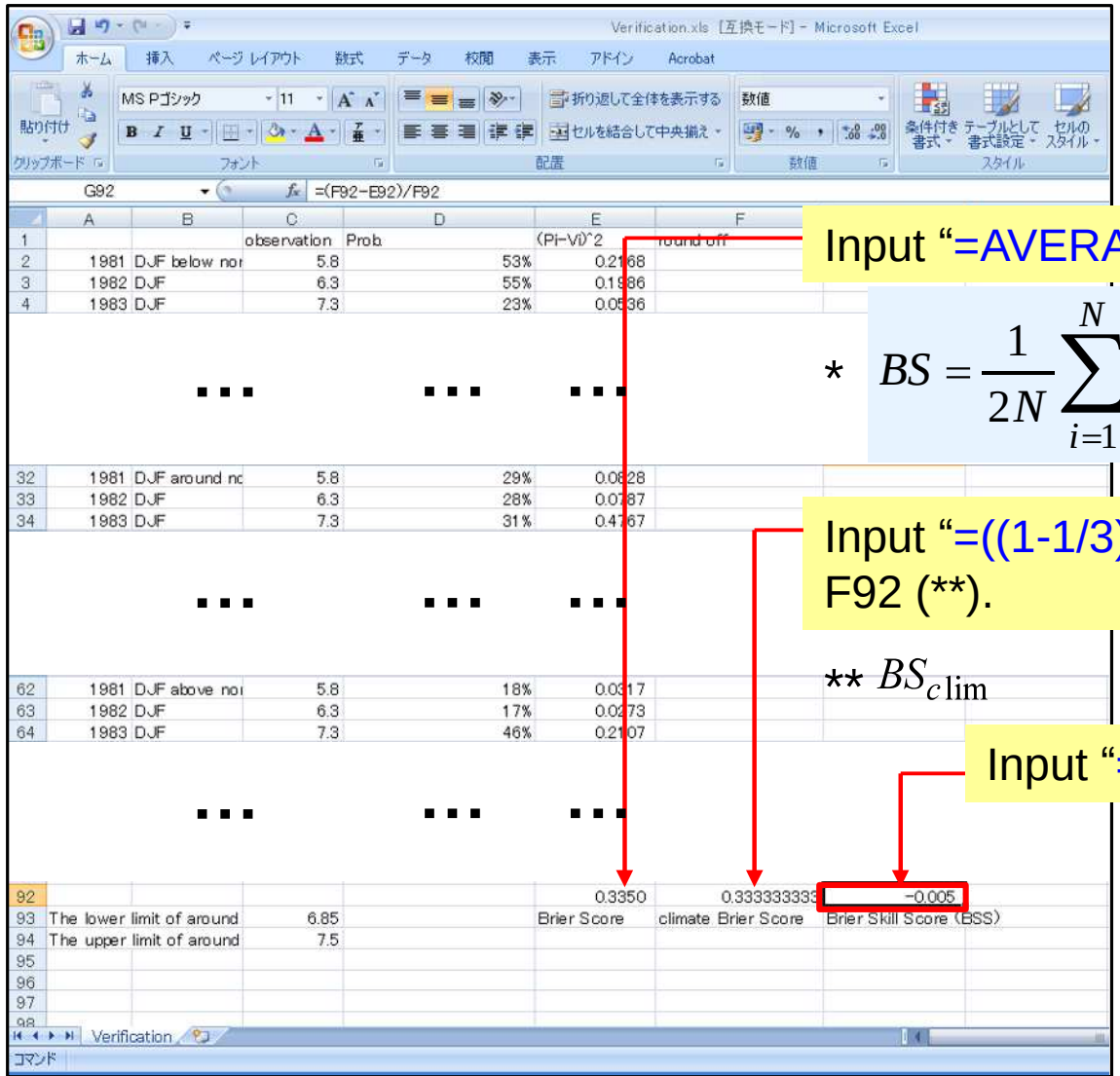
*** If the observed value is above normal, $E62 = (p_i{}^m - 1)^2$

If not so (below or around normal), $E62 = (p_i{}^m - 0)^2$

4-2. Probabilistic Verification

Brier skill score (3-category forecast)

* If you use BriefVerification.xls in place of Verification.xls, the procedure in this slide is unnecessary.



Input "=AVERAGE(E2:E91)*(3/2)" at E92 (*).

* $BS = \frac{1}{2N}\sum_{i=1}^{N}\sum_{m=1}^{3}(p_i^m - o_i^m)^2$

Input "=((1-1/3)^2+(0-1/3)^2+(0-1/3)^2)/2" at F92 (**).

** $BS_{c\lim}$

Input "=(F92-E92)/F92" at G92 (***).

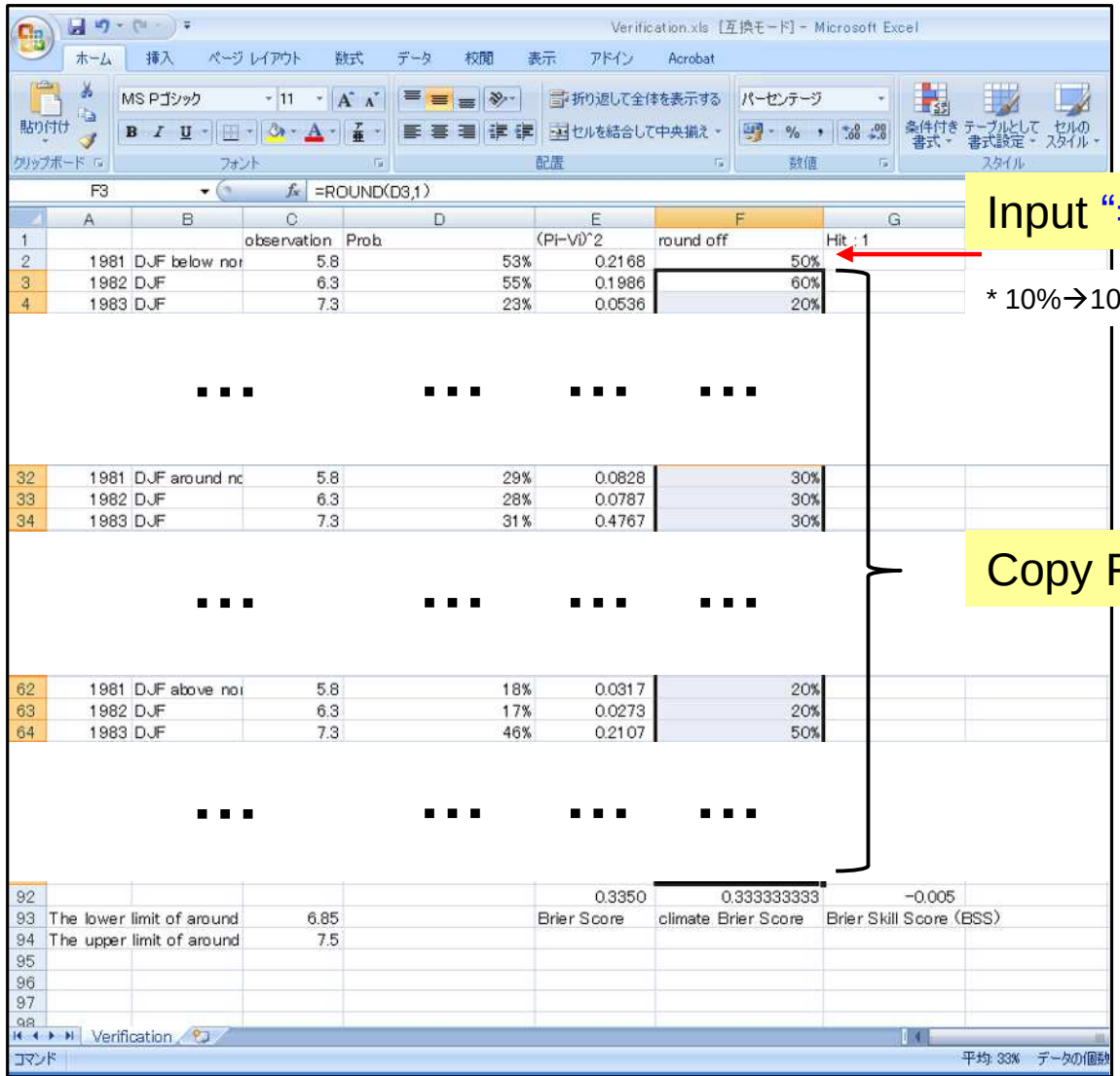
*** $BSS = \frac{BS - BS_{c\lim}}{0 - BS_{c\lim}}$

4-2. Probabilistic Verification

Reliability Diagram

•Continue using the Verification.xls.

＊ If you use BriefVerification.xls in place of Verification.xls, the procedure in this slide is unnecessary.



Input “=ROUND(D2,1)” at F2.

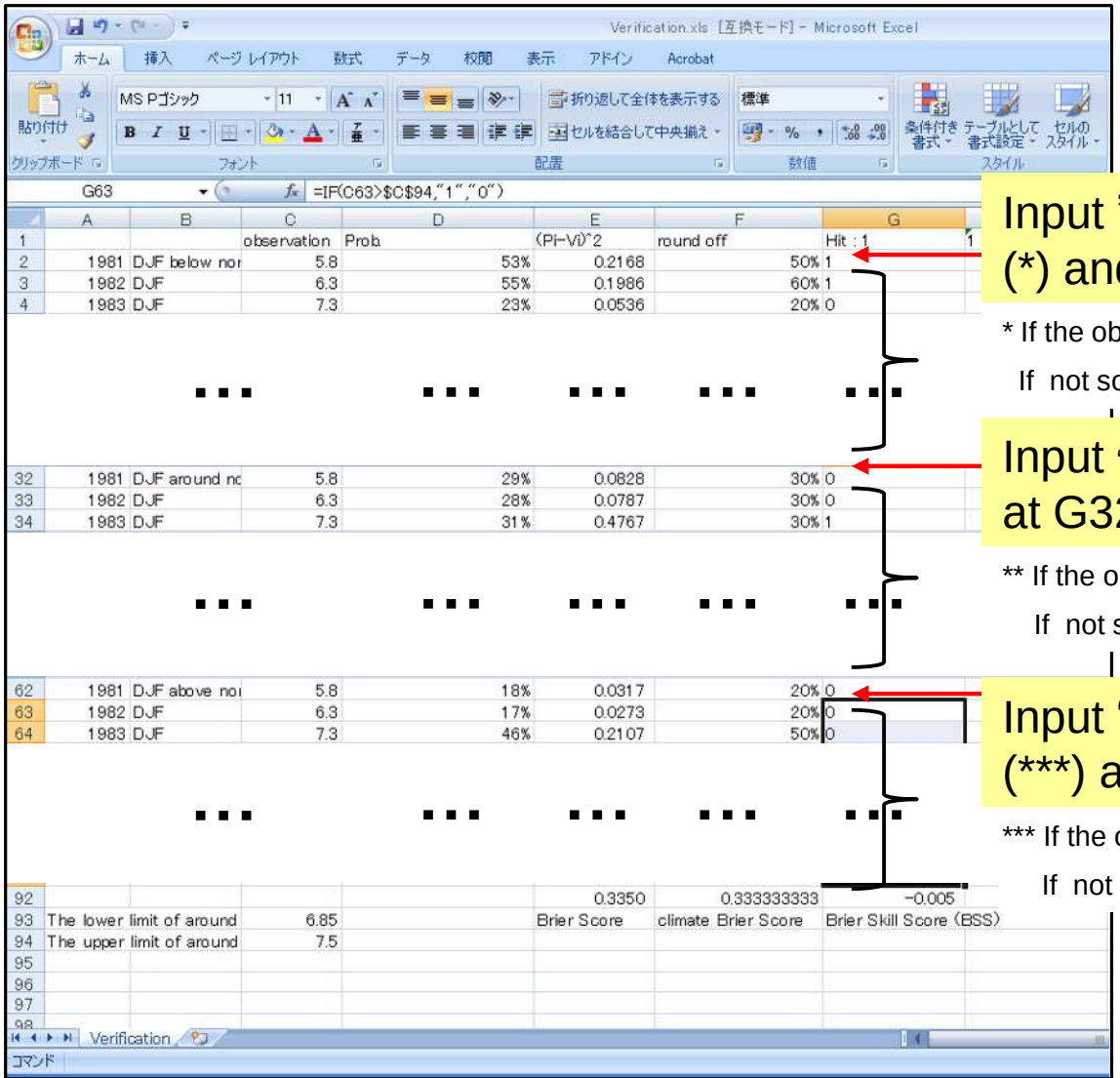
* 10%→10%, 7% →10%, 24% →20%, and so on.

Copy F2 and paste it intoF3:F91.

4-2. Probabilistic Verification

Reliability Diagram

* If you use BriefVerification.xls in place of Verification.xls, the procedure in this slide is unnecessary.



Input "=IF($C2<=$C$93,"1","0")" at G2 (*) and paste it into G3:G31.

* If the observed value is below normal, G2=1.
If not so (around or above normal), G2=0.

Input "=IF(C32<=$C$93,"0",IF(C32>$C$94,"0","1"))" at G32 (**) and paste it into G33:E61.

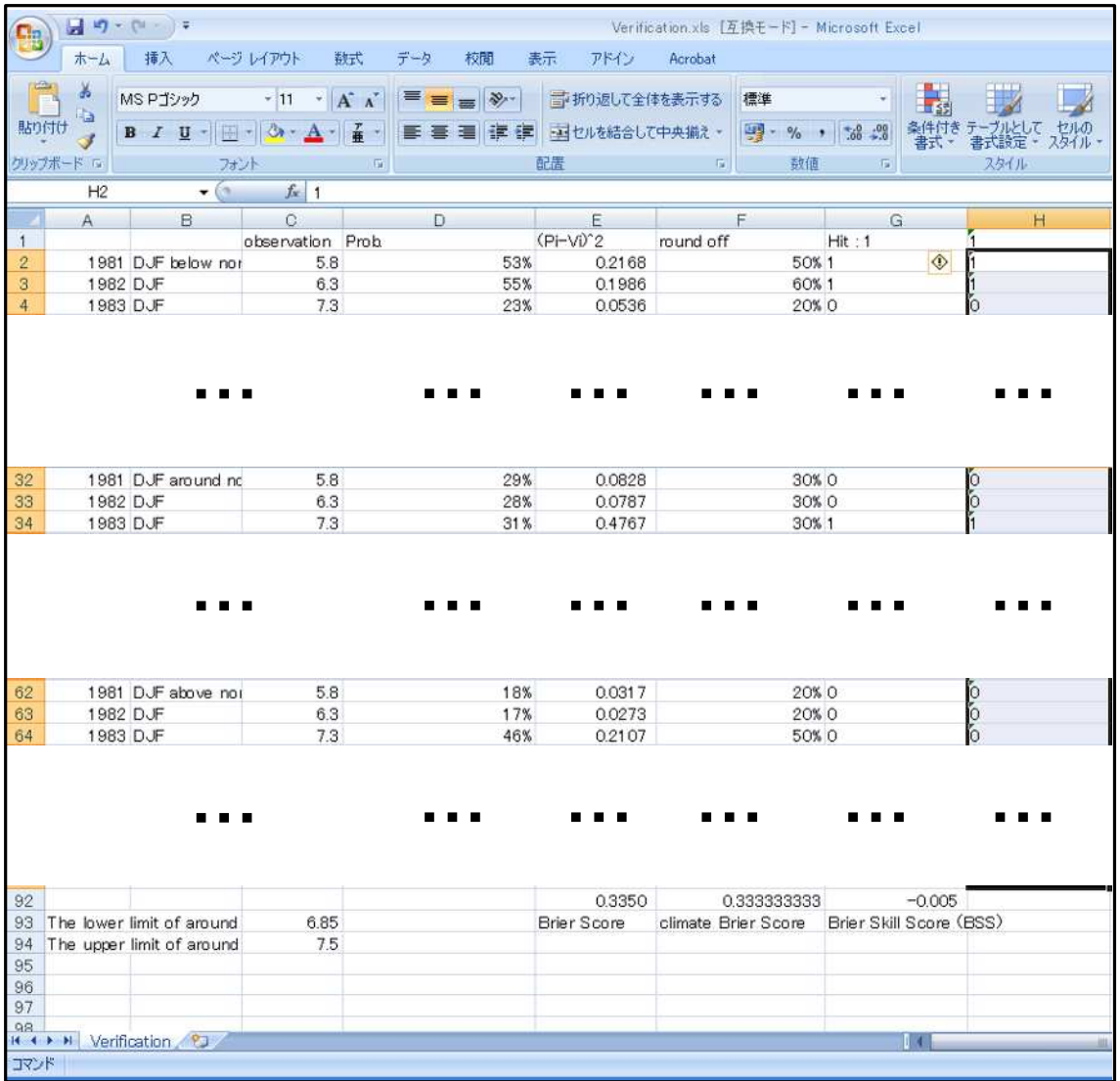
** If the observed value is below or above normal, G32=0.
If not so (around normal), G32=1.

Input "=IF(C62>$C$94,"1","0")" at G62 (***) and paste it into G63:G91.

*** If the observed value is above normal, G62=1.
If not so (below or around normal), G62=0.

4-2. Probabilistic Verification

Reliability Diagram



	A	B	C	D	E	F	G	H
1			observation	Prob.	(Pi-Vi)^2	round off	Hit : 1	1
2	1981	DJF below nor	5.8	53%	0.2168	50%	1	1
3	1982	DJF	6.3	55%	0.1986	60%	1	1
4	1983	DJF	7.3	23%	0.0536	20%	0	0
...								
32	1981	DJF around nc	5.8	29%	0.0828	30%	0	0
33	1982	DJF	6.3	28%	0.0787	30%	0	0
34	1983	DJF	7.3	31%	0.4767	30%	1	1
...								
62	1981	DJF above nor	5.8	18%	0.0317	20%	0	0
63	1982	DJF	6.3	17%	0.0273	20%	0	0
64	1983	DJF	7.3	46%	0.2107	50%	0	0
...								
92					0.3350	0.333333333	-0.005	
93	The lower limit of around		6.85		Brier Score	climate Brier Score	Brier Skill Score (BSS)	
94	The upper limit of around		7.5					

Copy G2:G91, and paste their values into H2:H91 (functions → values).

4-2. Probabilistic Verification

Reliability Diagram

(1) Frequency of forecast XXX%

=COUNTIF($F$2:$F$91,"XXX%")

* XXX: 0,10, …,100

(2) Hit frequency of forecast XXX%

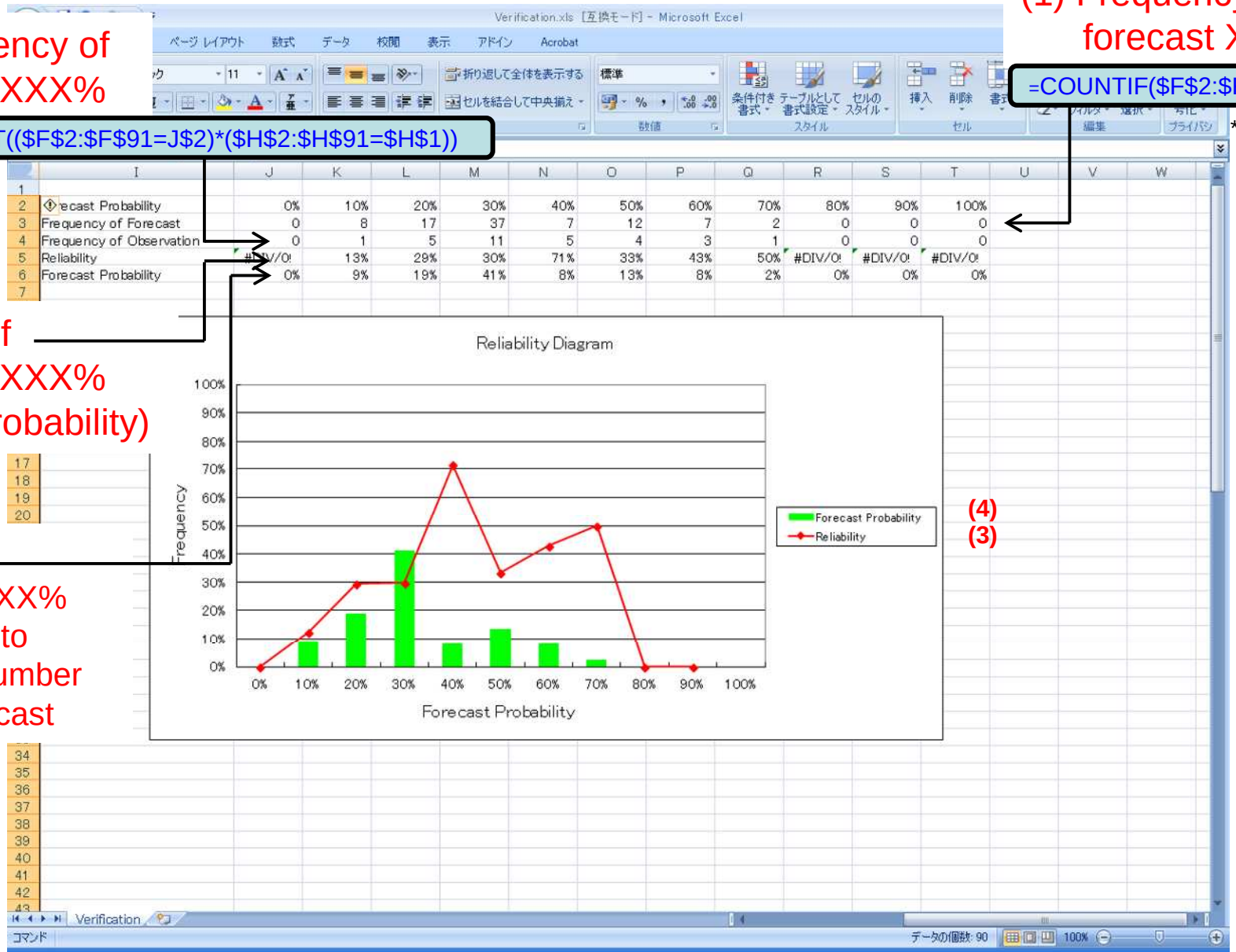
=SUMPRODUCT(($F$2:$F$91=J$2)*($H$2:$H$91=$H$1))

(3) Hit rate of forecast XXX% (observed probability)

=J4/J3

(4) Ratio of forecast XXX% compared to the total number of the forecast

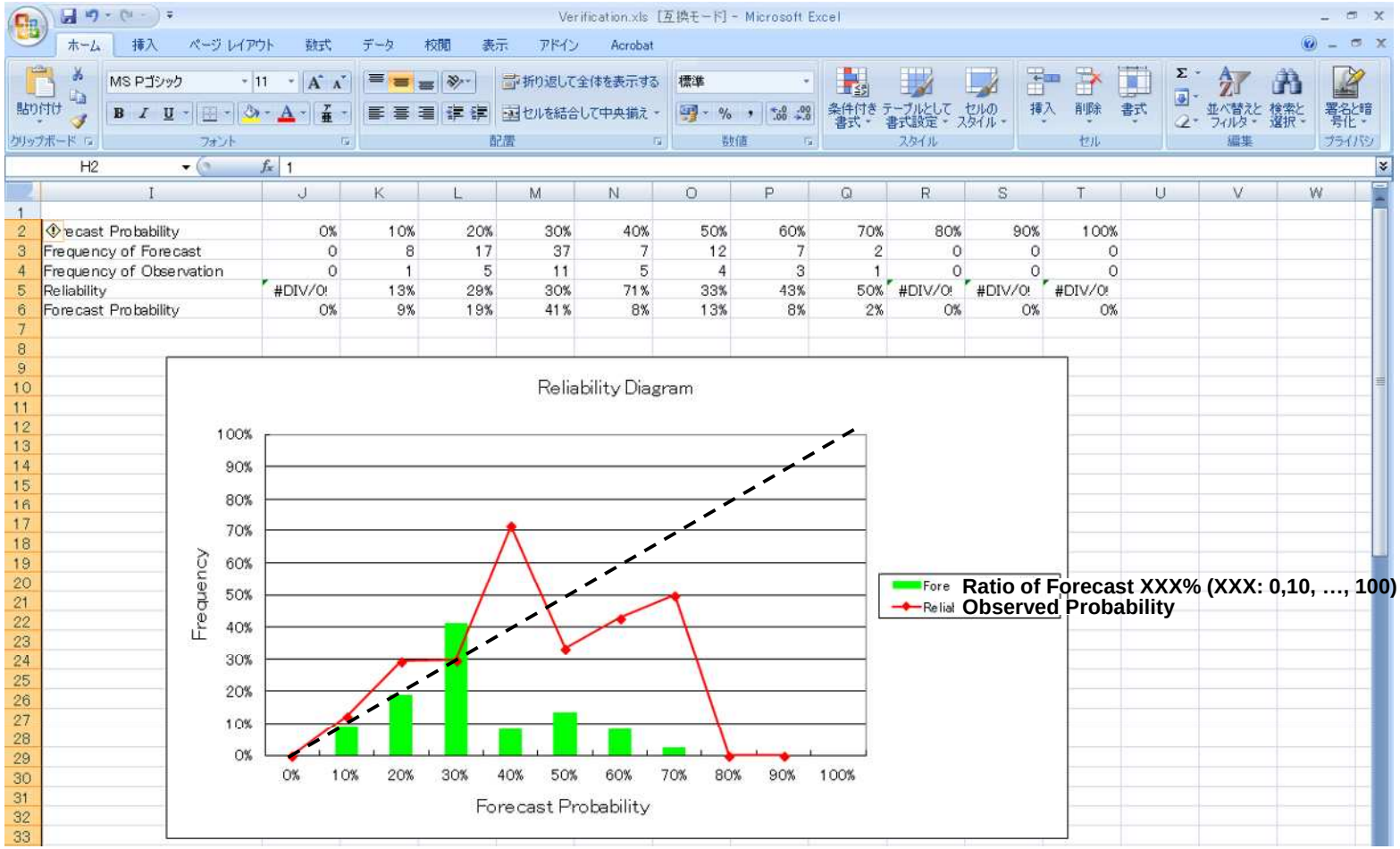
=J3/90



	I	J	K	L	M	N	O	P	Q	R	S	T	U
2	Forecast Probability	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	
3	Frequency of Forecast	0	8	17	37	7	12	7	2	0	0	0	
4	Frequency of Observation	0	1	5	11	5	4	3	1	0	0	0	
5	Reliability	#DIV/0!	13%	29%	30%	71%	33%	43%	50%	#DIV/0!	#DIV/0!	#DIV/0!	
6	Forecast Probability	0%	9%	19%	41%	8%	13%	8%	2%	0%	0%	0%	

4-2. Probabilistic Verification

Reliability Diagram



	I	J	K	L	M	N	O	P	Q	R	S	T
2	Forecast Probability	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
3	Frequency of Forecast	0	8	17	37	7	12	7	2	0	0	0
4	Frequency of Observation	0	1	5	11	5	4	3	1	0	0	0
5	Reliability	#DIV/0!	13%	29%	30%	71%	33%	43%	50%	#DIV/0!	#DIV/0!	#DIV/0!
6	Forecast Probability	0%	9%	19%	41%	8%	13%	8%	2%	0%	0%	0%

- The reliability diagram plots the observed probability(Y-axis) against the forecast probability(X-axis).
- The diagonal broken line indicates perfect reliability (observed probability equal to forecast probability).
- Points below (above) the diagonal line indicate overforecasting (underforecasting).

Questions

Single Regression Model

1. “What predictor did you select?”
2. “Could you get an accurate guidance?”

Multiple Regression Model

1. “What predictors did you select?
2. “Is the multiple regression model guidance more accurate than the single regression model guidance?”

Probabilistic Forecast

1. “How did the guidance predict the hottest/coldest/wettest/driest year in your country?”
2. “Are there any essential differences of the probability between the temperature and precipitation?”

Verification

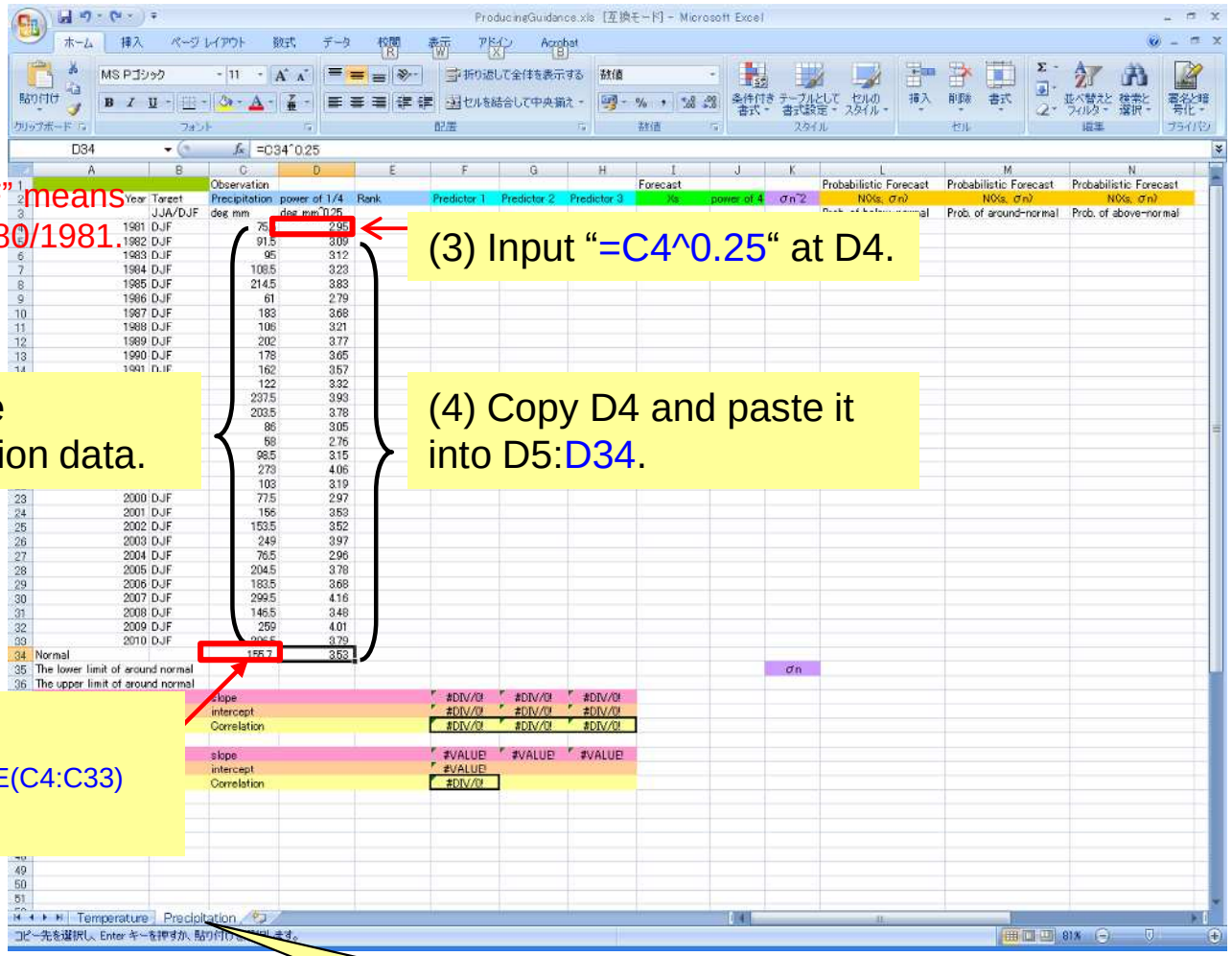
1. Is your guidance skillful from the viewpoint of deterministic forecast?
(Anomaly Correlation, Root Mean Square Error, Time Series Chart, and Scatter Plot)
2. Is your guidance skillful from the viewpoint of probabilistic forecast?
(Brier skill score and Reliability Diagram)

Please answer these questions as many as possible in your presentation time (17 minutes) after the Exercise Part 2.

Appendix

1. Single Regression Model

- Open the ProducingGuidance.xls.
- Paste observation data, calculate the normal from 1981 to 2010, and calculate the power of 1/4 of precipitation values.



"1981 DJF" means DJF of 1980/1981.

(1) Paste observation data.

(2) Input "=AVERAGE(C4:C33)" at C34.

(3) Input "=C4^0.25" at D4.

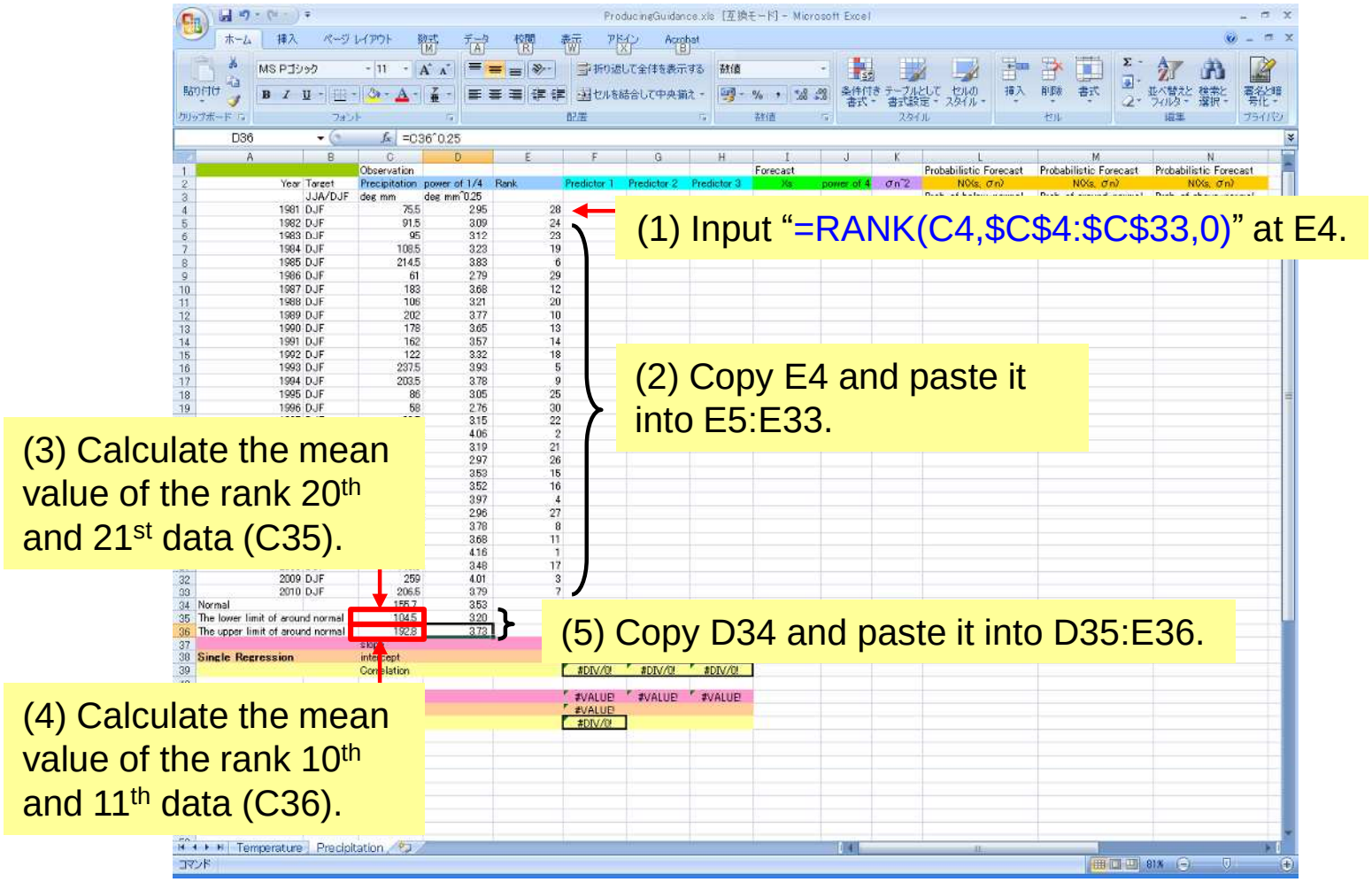
(4) Copy D4 and paste it into D5:D34.

Precipitation worksheet

In this case, the predictand is precipitation and the station is Tokyo.

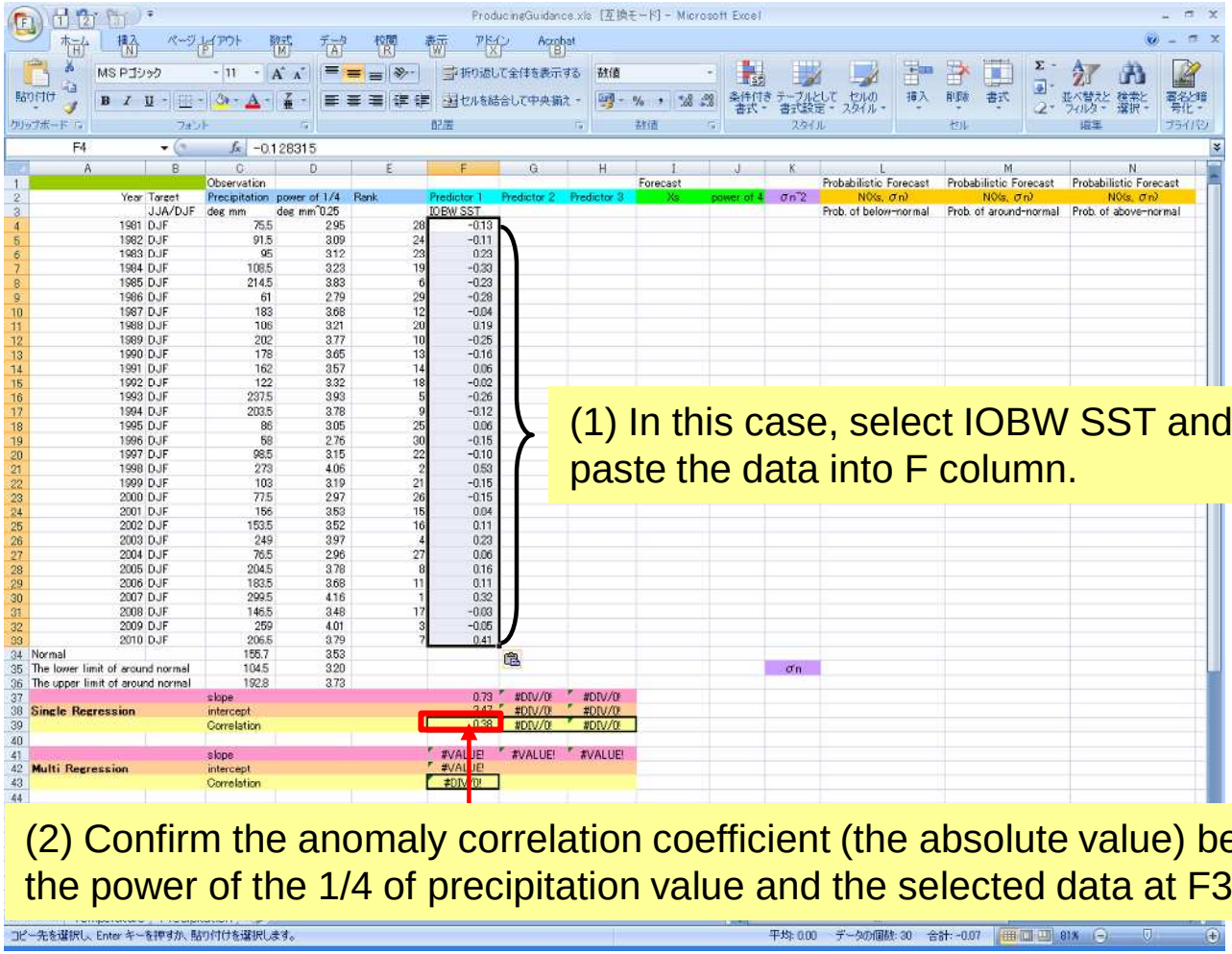
1. Single Regression Model

- Rank the data from 1981 to 2010 in descending order and calculate the lower and upper limits of around normal (preparation for probabilistic forecast).



1. Single Regression Model

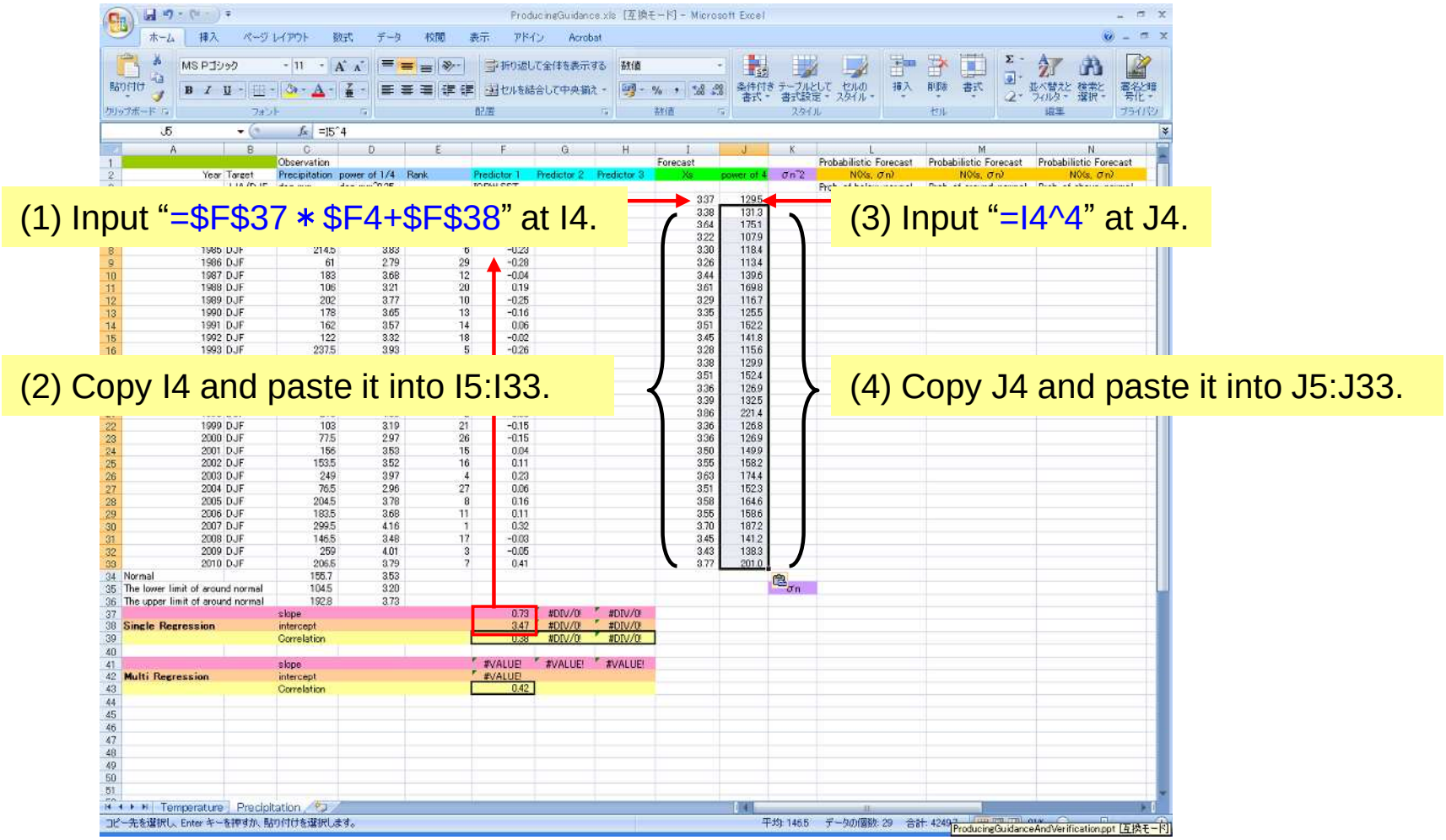
- Open GPVdata.xls and indices.xls.
- Select a predictor and paste the data into F column



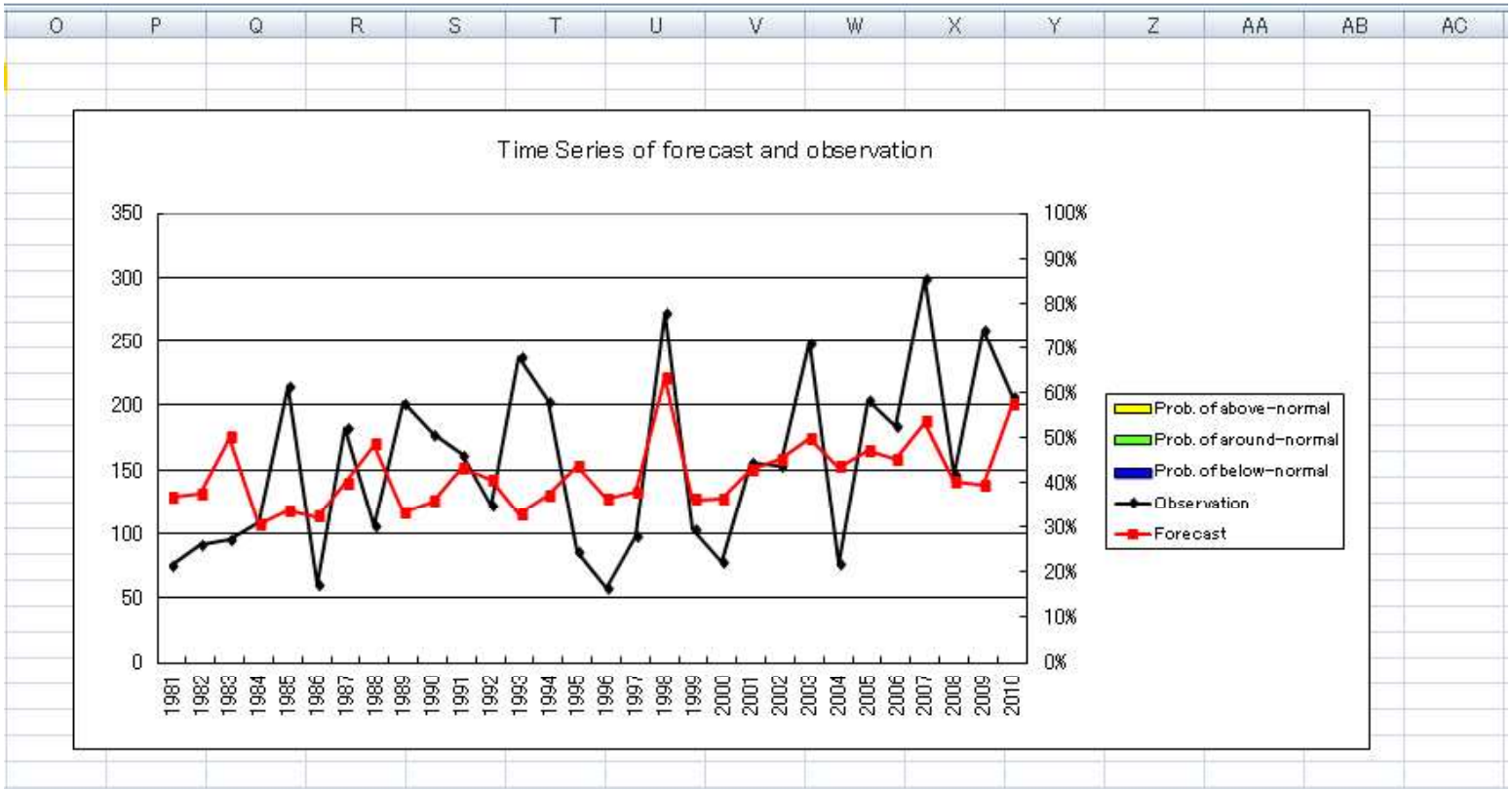
Please try to look for the most effective predictor.

1. Single Regression Model

- Calculate the power of 1/4 of the precipitation value using single regression equation.
- Calculate the 4th powers of them. (That is to say, forecasted precipitation values.)



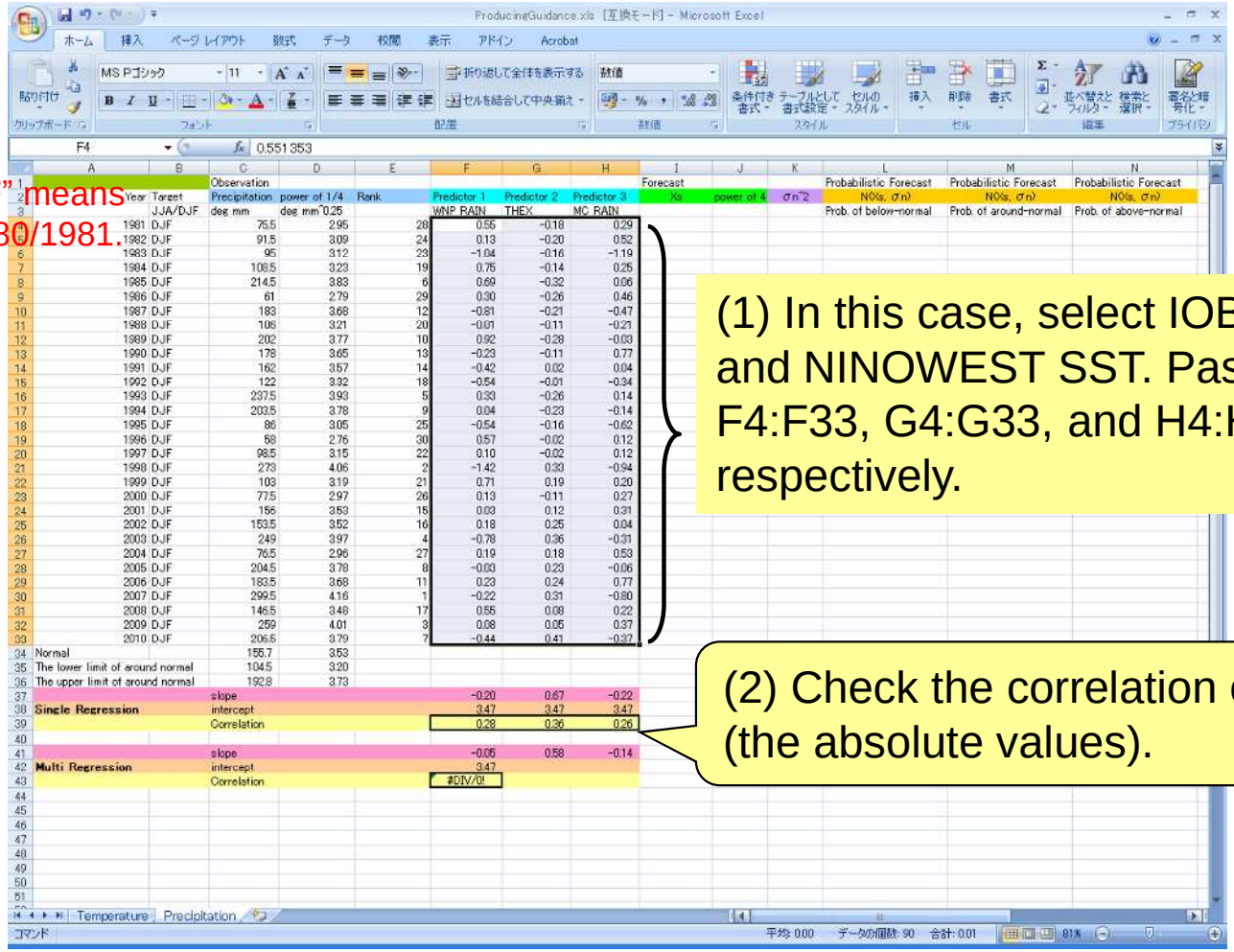
1. Single Regression Model



2. Multiple Regression Model

- Try to look for most effective combination of predictors. after trial and error. The ITACS will help you find them (see the slides 21-23).

"1981 DJF" means DJF of 1980/1981.

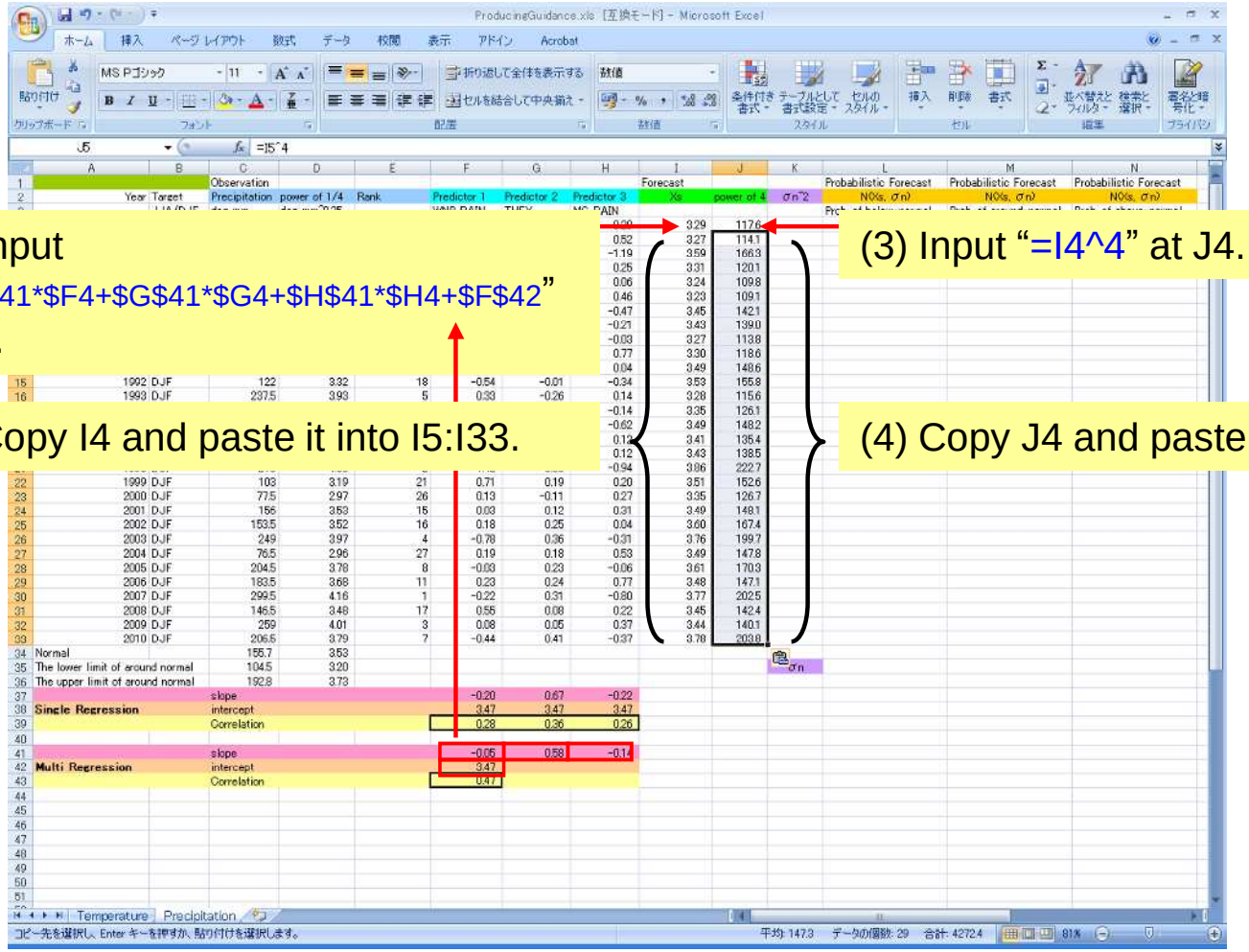


(1) In this case, select IOBW, Z2030, and NINOWEST SST. Paste them to F4:F33, G4:G33, and H4:H33, respectively.

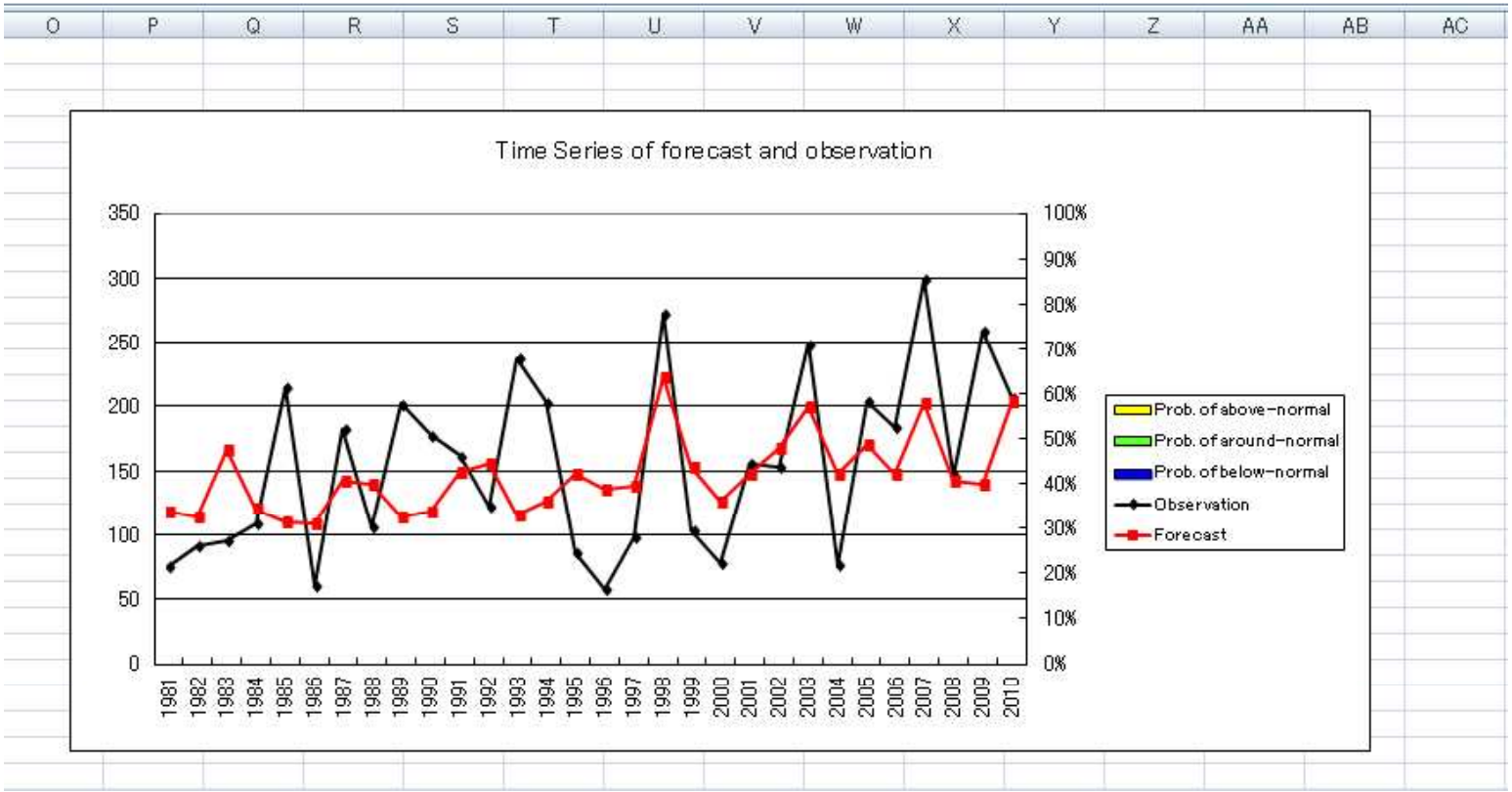
(2) Check the correlation coefficients (the absolute values).

2. Multiple Regression Model

- Calculate the power of 1/4 of the precipitation value using multiple regression equation.
- Calculate the 4th powers of them. (That is to say, forecasted precipitation values.)



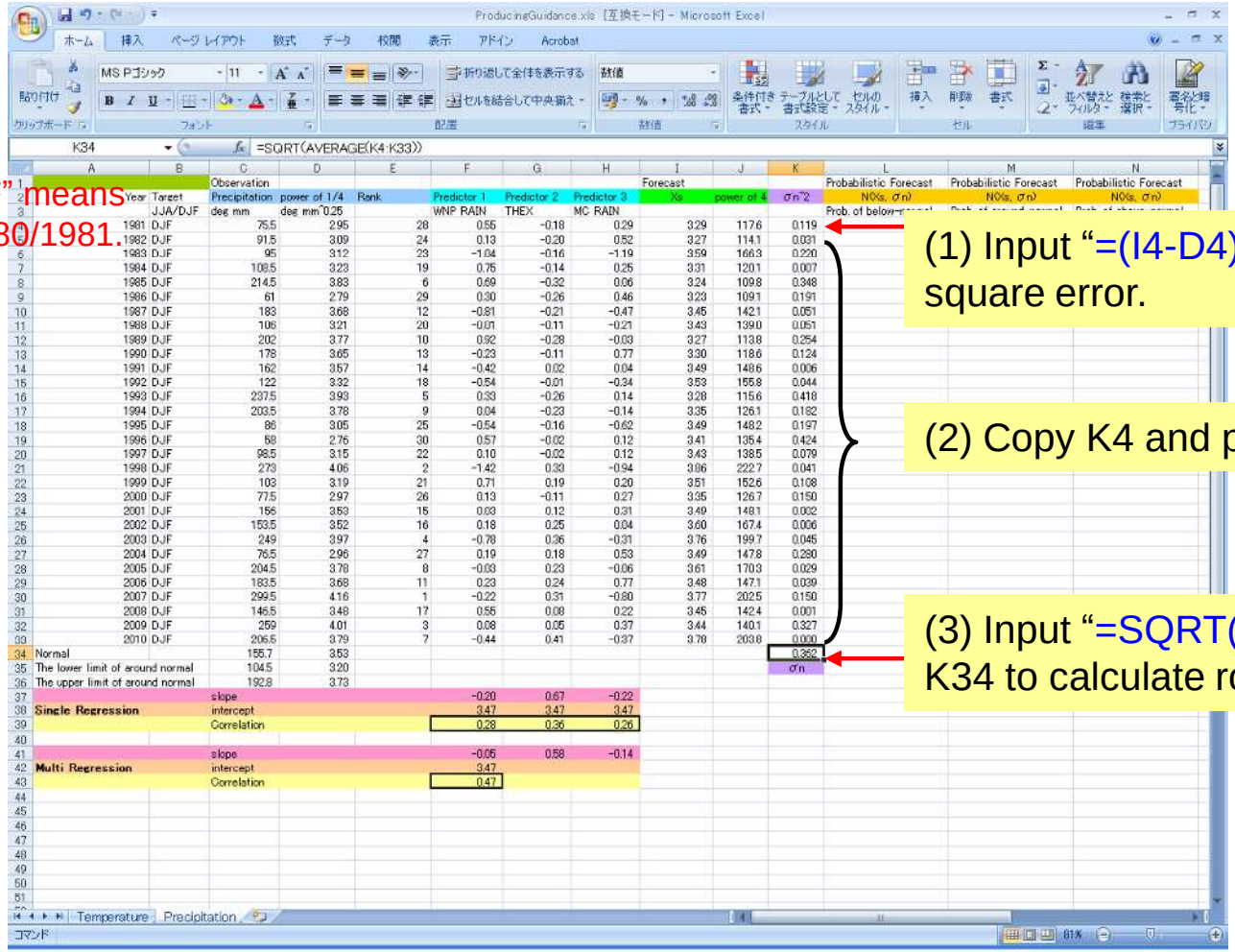
2. Multiple Regression Model



3. Probabilistic Forecast

- Calculate squares of regression errors of the powers of 1/4 of precipitation values .
- Calculate the root mean square error.

“1981 DJF” means DJF of 1980/1981.



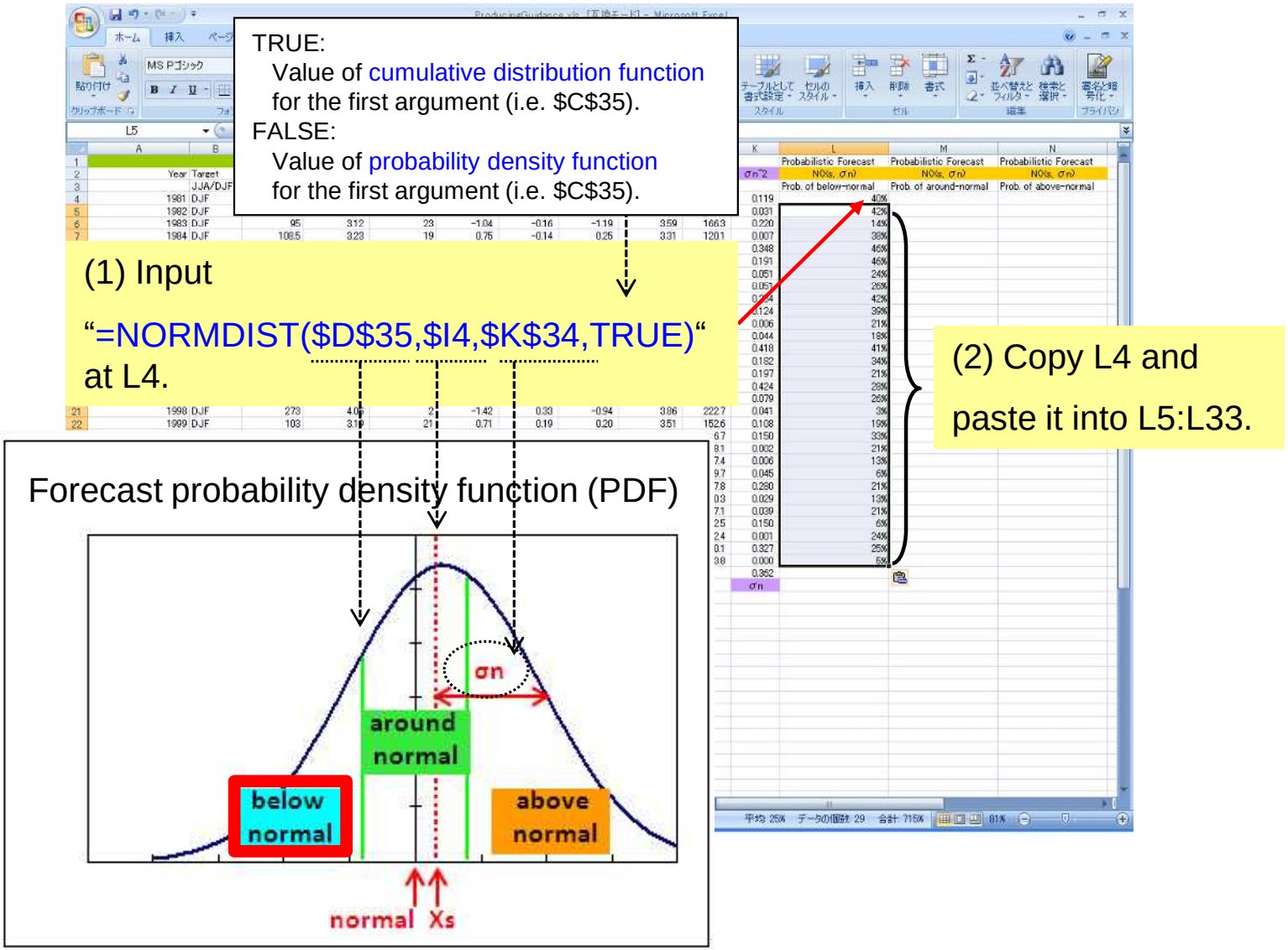
(1) Input “=(I4-D4)^2” at K4 to calculate the square error.

(2) Copy K4 and paste it into K5:K33.

(3) Input “=SQRT(AVERAGE(K4:K33))” at K34 to calculate root mean square error.

3. Probabilistic Forecast

- Calculate probability of below-normal.



3. Probabilistic Forecast

- Calculate probability of above-normal.

Very important!

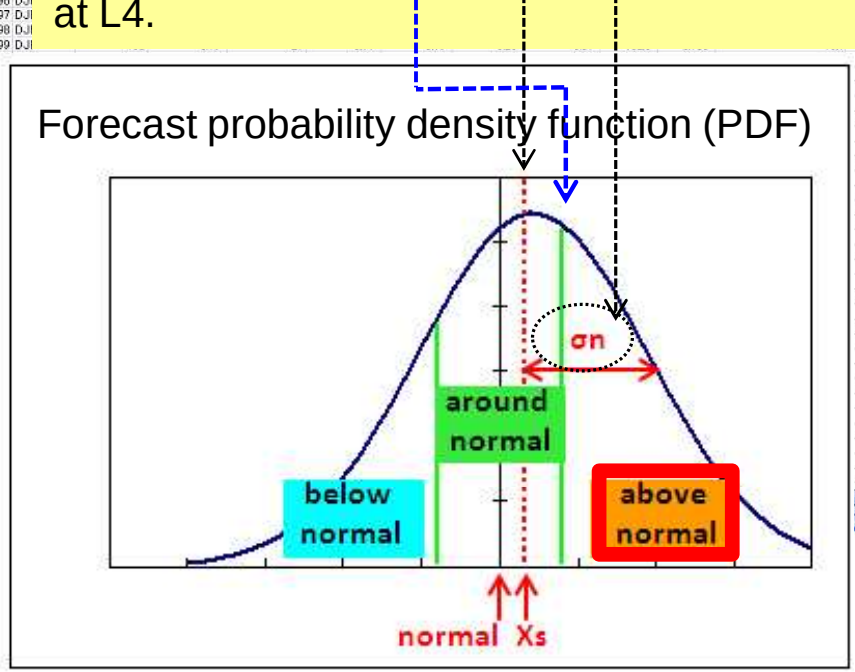
TRUE:
Value of cumulative distribution function for the first argument (i.e. $C$36).
FALSE:
Value of probability density function for the first argument (i.e. $C$36).

(1) Input

"=1-NORMDIST($D$36,$I4,$K$34,TRUE)"

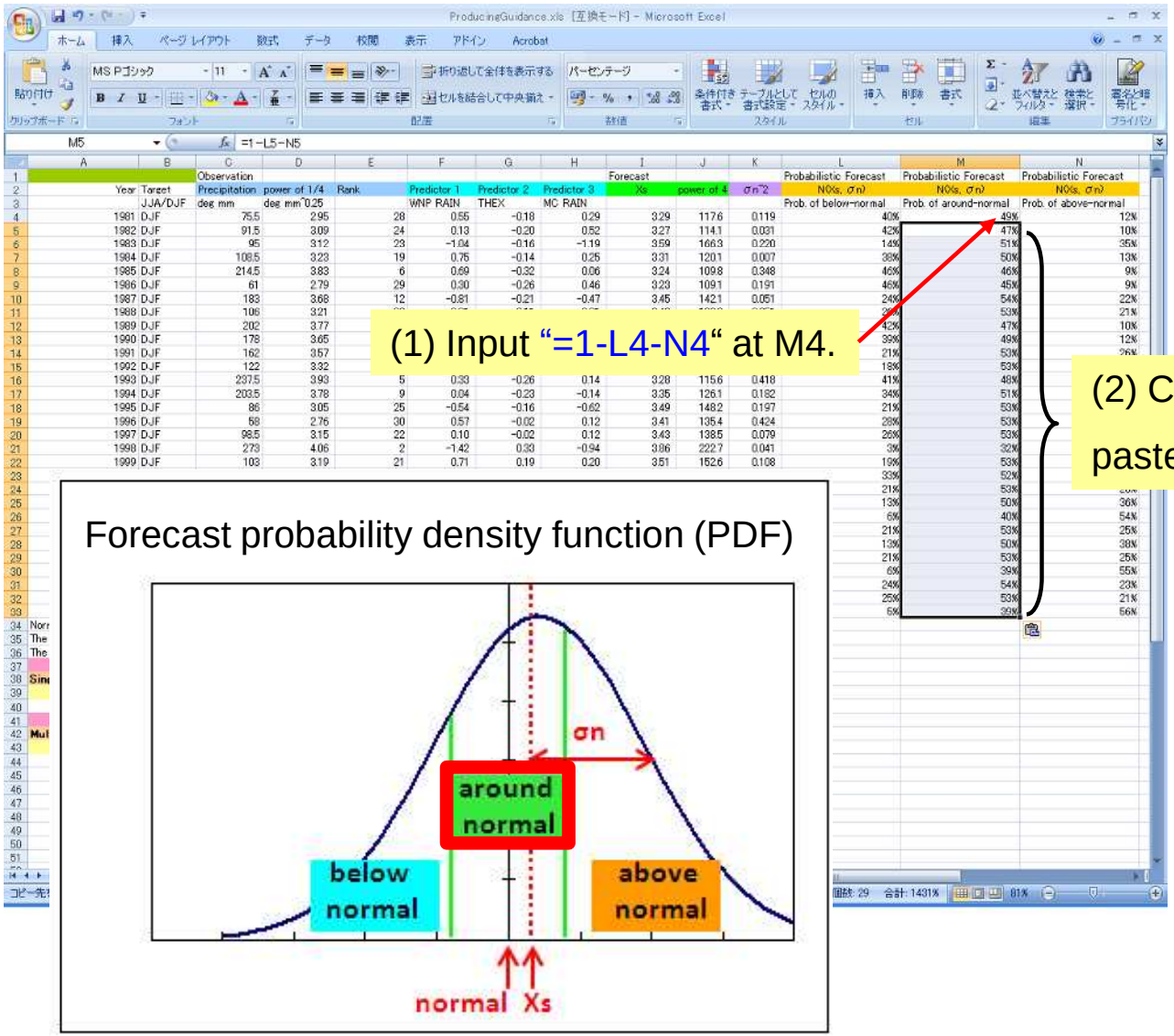
at L4.

(2) Copy N4 and paste it into N5:N33.

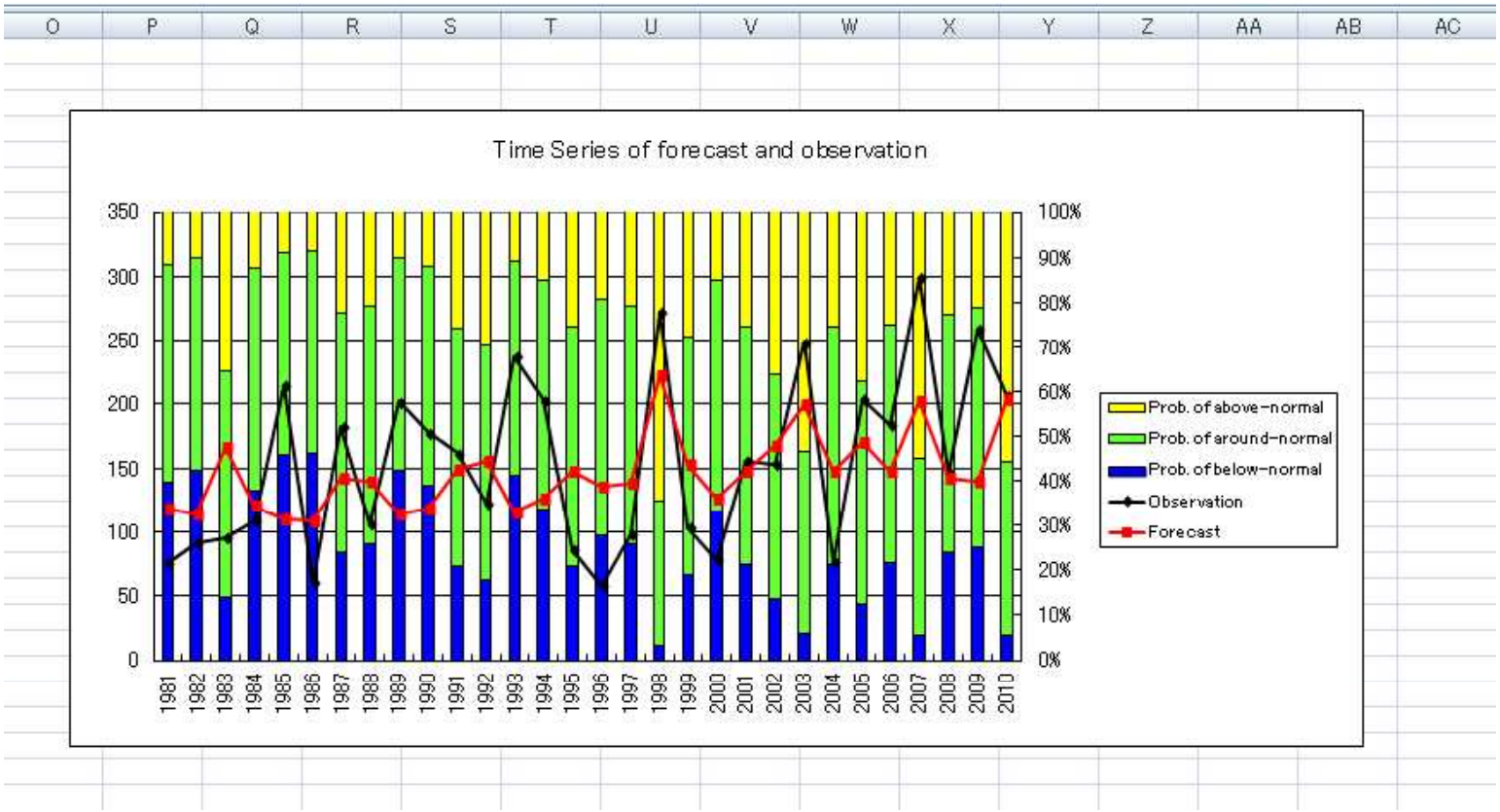


3. Probabilistic Forecast

- Calculate probability of around-normal.



3. Probabilistic Forecast



Anomaly Correlation (AC)

$$AC = \frac{\sum_{i=1}^{N}(F_i - C_i)(O_i - C_i)}{\sqrt{\sum_{i=1}^{N}(F_i - C_i)^2}\sqrt{\sum_{i=1}^{N}(O_i - C_i)^2}}$$

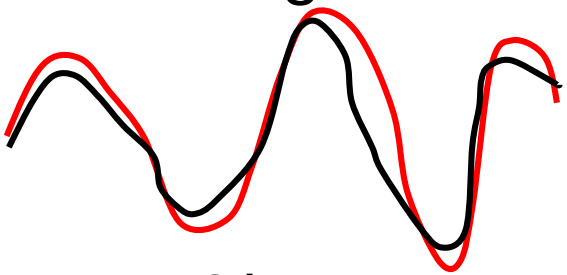
F :**forecast**

O :**observation**

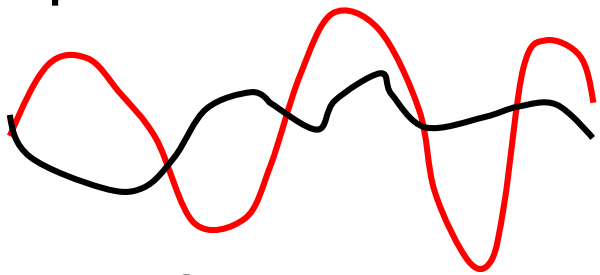
C :**climatology**

Range: -1 to 1. Perfect score: 1.

- AC measures correspondence or phase difference between forecast and observation, subtracting out the climatological mean at each point.



AC≒1



AC <<1

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(F_i - O_i\right)^2}$$

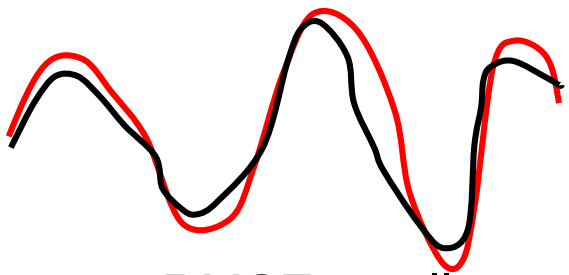
F : **forecast**

O : **observation**

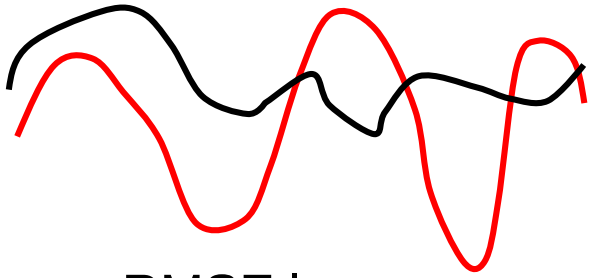
N : **sample size**

Range: 0 to infinity, Perfect score: 0.

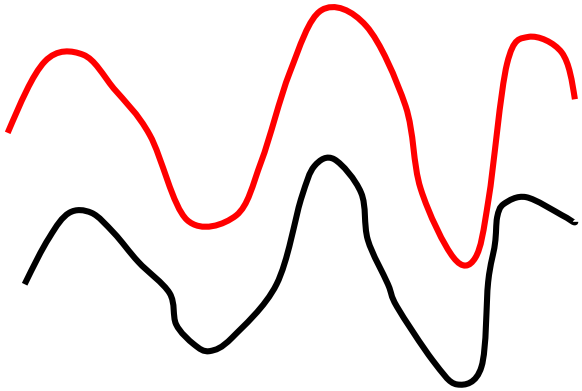
- RMSE measures absolute magnitude of the forecast error.
- It does not indicate the direction the error.



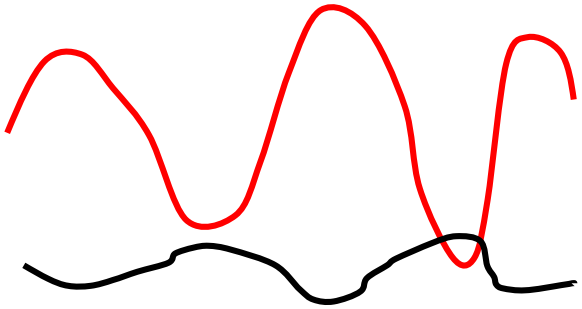
RMSE small



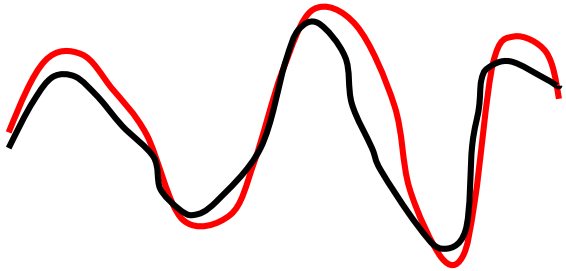
RMSE large



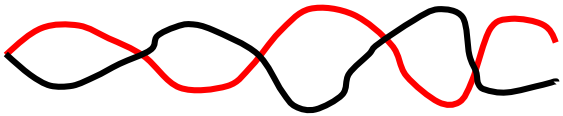
AC positive (≒1)
RMSE large



AC negative
RMSE large



AC positive(≒1)
RMSE small



AC negative
RMSE small