

ISA 444: Business Forecasting

05 - Basic Tools and Goodness of Fit (Cont.)

Fadel M. Megahed

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: [Click here to schedule an appointment](#)

Fall 2020

Outline

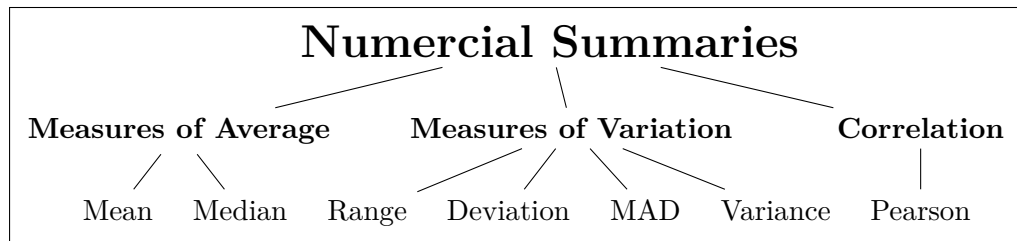
- 1 **Preface**
- 2 Transformations
- 3 Measures of Forecast Accuracy
- 4 Prediction Intervals
- 5 Recap

Quick Refresher based on Last Class

Main Learning Outcomes

- ✓ Use numerical summaries to describe a time series.
- ✗ Apply transformations to a time series.

Recap: Numerical Summaries



An overview of the numerical summaries discussed last class.

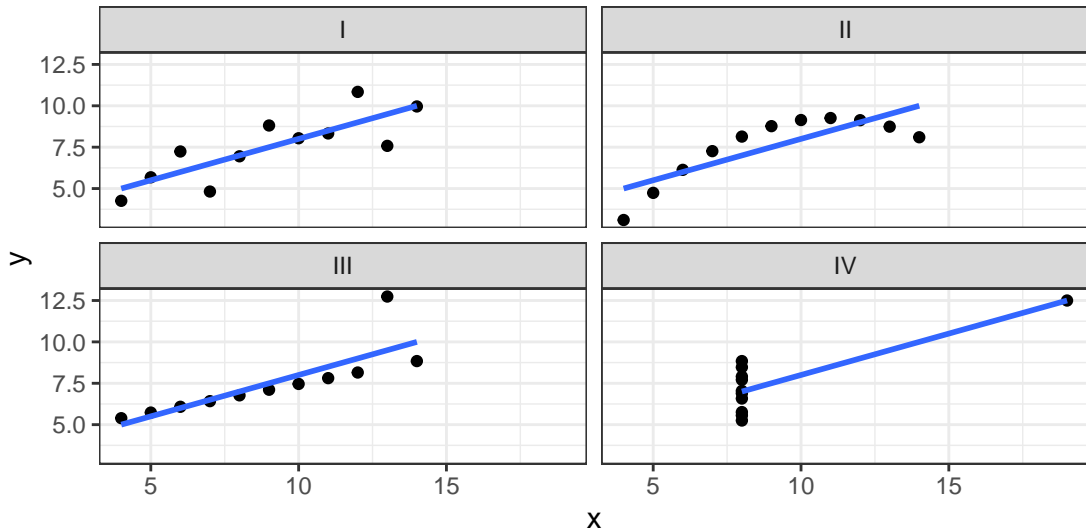
Recap: Numerical Summaries do NOT Replace Visuals [1]

x1	x2	x3	x4	y1	y2	y3	y4
10.00	10.00	10.00	8.00	8.04	9.14	7.46	6.58
8.00	8.00	8.00	8.00	6.95	8.14	6.77	5.76
13.00	13.00	13.00	8.00	7.58	8.74	12.74	7.71
9.00	9.00	9.00	8.00	8.81	8.77	7.11	8.84
11.00	11.00	11.00	8.00	8.33	9.26	7.81	8.47
14.00	14.00	14.00	8.00	9.96	8.10	8.84	7.04
6.00	6.00	6.00	8.00	7.24	6.13	6.08	5.25
4.00	4.00	4.00	19.00	4.26	3.10	5.39	12.50
12.00	12.00	12.00	8.00	10.84	9.13	8.15	5.56
7.00	7.00	7.00	8.00	4.82	7.26	6.42	7.91
5.00	5.00	5.00	8.00	5.68	4.74	5.73	6.89

Recap: Numerical Summaries do NOT Replace Visuals [2]

set	x.mean	x.sd	y.mean	y.sd	corr
I	9.00	3.32	7.50	2.03	0.82
II	9.00	3.32	7.50	2.03	0.82
III	9.00	3.32	7.50	2.03	0.82
IV	9.00	3.32	7.50	2.03	0.82

Recap: Numerical Summaries do NOT Replace Visuals [3]



Recap: First Differences [1]

The change in the time series from one period to the next is known as the (first) difference. It can be computed as follows:

$$DY_t = Y_t - Y_{t-1} \quad (1)$$

```
dowJonesIndex = tq_get("^DJI", from = "2020-08-17", to = "2020-08-22") %>%
  select(symbol, date, adjusted)
dowJonesIndex$`Yt-1` = lag(dowJonesIndex$adjusted)
dowJonesIndex$DYt = dowJonesIndex$adjusted - dowJonesIndex$`Yt-1`
```

symbol	date	adjusted	Yt-1	DYt
^DJI	2020-08-17	27844.91		
^DJI	2020-08-18	27778.07	27844.91	-66.84
^DJI	2020-08-19	27692.88	27778.07	-85.19
^DJI	2020-08-20	27739.73	27692.88	46.85
^DJI	2020-08-21	27930.33	27739.73	190.60

Recap: First Differences [2]

Note that the differences can be computed in one step using the function `diff()` from base R as follows.

```
dowJonesIndex = tq_get("^DJI", from = "2020-08-17", to = "2020-08-22") %>%
  select(symbol, date, adjusted)
dowJonesIndex$DYt = c(NA, diff(dowJonesIndex$adjusted))
```

symbol	date	adjusted	DYt
^DJI	2020-08-17	27844.91	
^DJI	2020-08-18	27778.07	-66.84
^DJI	2020-08-19	27692.88	-85.19
^DJI	2020-08-20	27739.73	46.85
^DJI	2020-08-21	27930.33	190.60

Recap: Growth Rates in Practice – a Non-Graded Class Activity

- **Follow up question:** Hypothetically speaking let us say that Fadel has purchased \$10 worth of [Stellar \(\\$XLM\)](#). Each \$XLM coin was worth 0.1 on 08-25-2020.
 - Let us say that Fadel was lucky and his investment went up 20% on 08-26-2020, i.e. $GY_t = 20$. Using the formula, compute the value of each coin on 08-26-2020.
 - Due to the volatility of cryptocurrencies, let us assume that Fadel's growth rate for the following day was -20% (i.e., in comparison with the coin's value on 08-26-2020). Compute the value of the coin on 08-27-2020.
 - **Recall:**
 - *Value of coin on 08-26-2020: 0.12.*
 - *Value of coin on 08-27-2020: 0.096.*

Learning Objectives for Today's Class

Main Learning Outcomes

- Apply (log) transformations to a time series.
- Apply and interpret measures of forecast accuracy.
- Interpret prediction intervals for a simple forecast.

Outline

- 1 Preface
- 2 Transformations**
- 3 Measures of Forecast Accuracy
- 4 Prediction Intervals
- 5 Recap

The Log Transform [1]

The log transformation can be computed as follows:

$$L_t = \ln(Y_t) \quad (2)$$

Note that the `log()` in R takes the natural logarithm as its default base, i.e., would transform a variable/statistic based on the above equation.

The reverse transformation using the exponential function is:

$$e^{L_t} = e^{\ln(Y_t)} = Y_t \quad (3)$$

The first difference in logarithms represents the logarithm of the ratio:

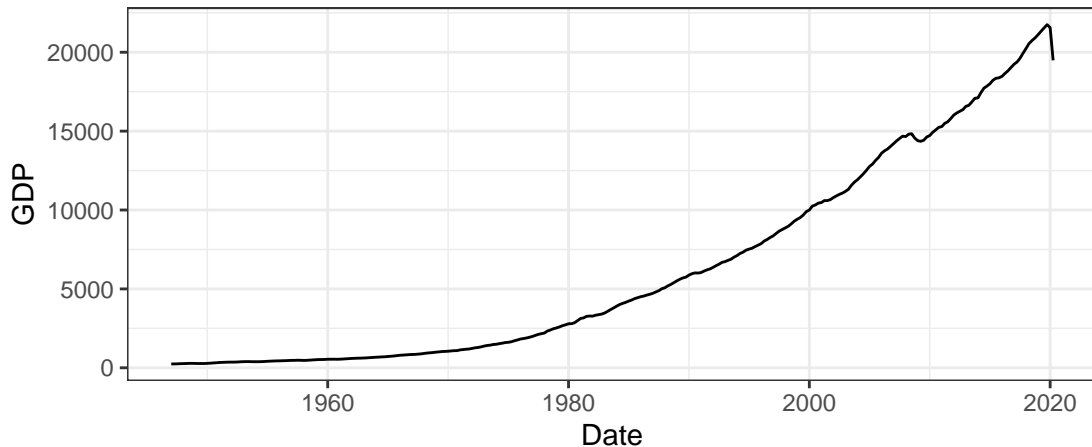
$$L_t = \ln\left(\frac{Y_t}{Y_{t-1}}\right) = \ln(Y_t) - \ln(Y_{t-1}) \quad (4)$$

The Log Transform [2]

- The primary purpose of the log transform is to **convert exponential growth into linear growth**.
- The transform often has the **secondary purpose of balancing the variance**.
- Difference in logs and growth rate transformations produce similar results and interpretations.

Plots with and without the Log Transformation [1]

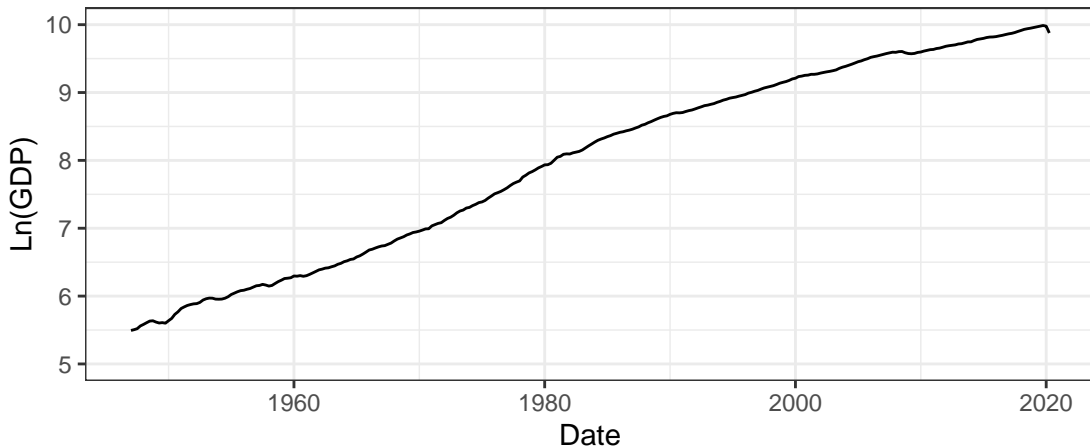
GDP without Log Transformation



Data from FRED

Plots with and without the Log Transformation [2]

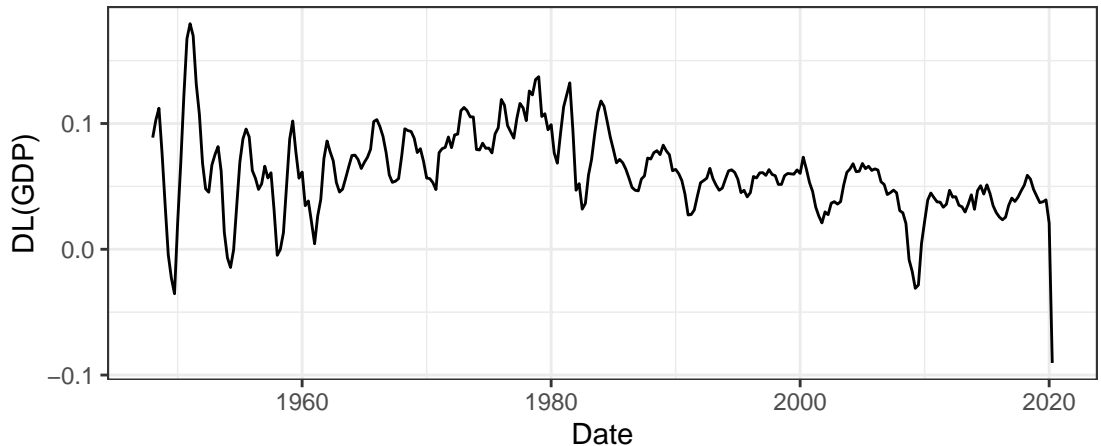
GDP with Log Transformation



Data from FRED

Plots with and without the Log Transformation [3]

Differences, with lag = 4, of the Log GDP

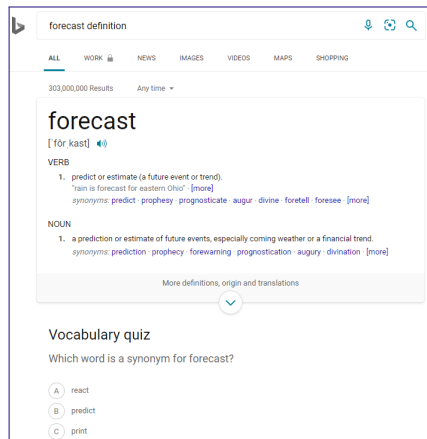


Data from FRED

Outline

- 1 Preface
- 2 Transformations
- 3 Measures of Forecast Accuracy**
- 4 Prediction Intervals
- 5 Recap

Recap: Definition of Forecast



The screenshot shows a Bing search result for the query "forecast definition". The search bar at the top contains the text "forecast definition" and has icons for voice search, image search, and a magnifying glass. Below the search bar are tabs for "ALL", "WORK", "NEWS", "IMAGES", "VIDEOS", "MAPS", and "SHOPPING". The "ALL" tab is selected. Below the tabs, it says "303,000,000 Results" and "Any time". The main content area displays the word "forecast" in a large font, followed by its phonetic transcription "[fɔrˌkɑst]" and a speaker icon. Below this, it lists the word as a "VERB" and a "NOUN". For the verb, the first definition is "predict or estimate (a future event or trend)", with an example "rain is forecast for eastern Ohio" and a link to "[more]". Synonyms listed are "predict", "prophesy", "prognosticate", "augur", "divine", "foretell", and "foresee", each with a "[more]" link. For the noun, the first definition is "a prediction or estimate of future events, especially coming weather or a financial trend", with synonyms "prediction", "prophecy", "forewarning", "prognostication", "augury", "divination", and "[more]". At the bottom of the definition section is a link "More definitions, origin and translations" with a downward arrow icon. Below the definition section is a "Vocabulary quiz" section with the question "Which word is a synonym for forecast?". There are three radio button options: A react, B predict, and C print. Option B is selected.

forecast definition

ALL WORK NEWS IMAGES VIDEOS MAPS SHOPPING

303,000,000 Results Any time

forecast
[fɔrˌkɑst]

VERB

1. predict or estimate (a future event or trend).
"rain is forecast for eastern Ohio" · [more]
synonyms: predict · prophesy · prognosticate · augur · divine · foretell · foresee · [more]

NOUN

1. a prediction or estimate of future events, especially coming weather or a financial trend.
synonyms: prediction · prophecy · forewarning · prognostication · augury · divination · [more]

More definitions, origin and translations

Vocabulary quiz

Which word is a synonym for forecast?

A react
B predict
C print

The definition of the term "forecast" as obtained from Bing/Merriam-Webster.

A Naïve Forecast

- A naïve forecast for an observation, Y_t , is the observation prior, Y_{t-1} .
- For some types of time series (e.g. Random Walks), a naïve forecast is the best possible forecast one can make¹.
- In the case of seasonal data, a naïve forecast could be the observation from the prior period.
 - For example, in the case of monthly data, the naïve forecast for the observation $Y_{Jan2018}$ could be $Y_{Jan2017}$. In this case, we would denote the frequency, $m=12$, and the naïve forecast for Y_t is the observation m periods prior, or Y_{t-m} .

¹Slide is from [Dr. Allison Jones-Farmer's](#) lecture notes, Miami University, Spring 2020.

Measures of Forecast Accuracy

The measures of accuracy we will discuss all deal with the difference between the actual observed value (Y_t) and the forecasted value (F_t) at time t . In order to measure forecast accuracy, we assume we have m actual values available, thus we have $Y_{t+1}, Y_{t+2}, \dots, Y_{t+m}$ and forecasts $F_{t+1}, F_{t+2}, \dots, F_{t+m}$. This is important because we will be averaging the forecast errors over m .²

Forecast Error:

$$e_{t+i} = Y_{t+i} - F_{t+i}. \quad (5)$$

²Note that your textbook discusses rolling forecast origins. This is important, but we will save this discussion for later in the semester. For now, assume the forecast origin is fixed (i.e., we are only interested in the one-period ahead forecast).

Measures Reflecting “Average” Forecast Performance

Mean Error:

$$ME = \frac{\sum_{i=1}^m e_{t+i}}{m}. \quad (6)$$

Mean Percentage Error:

$$MPE = \frac{100}{m} \sum_{i=1}^m \frac{e_{t+i}}{Y_{t+i}}. \quad (7)$$

Computing Measures of “Average” Forecast Performance [1]

symbol	date	adjusted	naiveFC	e	PE
AAPL	2020-08-17	458.43			
AAPL	2020-08-18	462.25	458.43	3.82	0.83
AAPL	2020-08-19	462.83	462.25	0.58	0.13
AAPL	2020-08-20	473.10	462.83	10.27	2.17
AAPL	2020-08-21	497.48	473.10	24.38	4.90
AAPL	2020-08-24	503.43	497.48	5.95	1.18
AAPL	2020-08-25	499.30	503.43	-4.13	-0.83
AAPL	2020-08-26	506.09	499.30	6.79	1.34
AAPL	2020-08-27	500.04	506.09	-6.05	-1.21
AAPL	2020-08-28	499.23	500.04	-0.81	-0.16

Based on Approach #1, the ME and MPE are equal to 4.53 and 0.93% , respectively.

Comments:

- We are forecasting the adjusted closing price of the AAPL stock for the next trading day, i.e., we are ignoring non-trading days.
- The naïve forecast is the `lag(1)` of the series; thus, the forecast error is the `diff(1)`.

Computing Measures of “Average” Forecast Performance [2]

```
pacman::p_load(tidyquant, magrittr, fpp2)

aapl = tq_get("AAPL", from = "2020-08-17", to = "2020-08-29") %>%
  select(symbol, date, adjusted)
naiveFC = snaive(aapl$adjusted) %>% .[['fitted']] #snaive from fpp2

e = aapl$adjusted - naiveFC
ME = mean(e, na.rm = T)
MPE = 100*mean(e/aapl$adjusted, na.rm = T)

cat(paste0("Based on Approach #2, the ME and MPE are equal to ", round(ME, 2)
          "%, respectively."))
```

Based on Approach #2, the ME and MPE are equal to 4.53 and 0.93%, respectively.

Measures Reflecting “Variability” in Forecast Performance

Absolute Forecast Error:

$$|e_{t+i}| = |Y_{t+i} - F_{t+i}|. \quad (8)$$

Squared Forecast Error:

$$(e_{t+i})^2 = (Y_{t+i} - F_{t+i})^2. \quad (9)$$

Mean Absolute Error:

$$MAE = \frac{\sum_{i=1}^m |e_{t+i}|}{m}. \quad (10)$$

Root Mean Squared Error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (e_{t+i}^2)}{m}}. \quad (11)$$

Measures Reflecting “Relative” Forecast Performance

Mean Absolute Percentage Error:

$$MAPE = \frac{100}{m} \sum_{i=1}^m \frac{|e_{t+i}|}{m}. \quad (12)$$

Relative Mean Absolute Error:

$$RelMAE = \frac{\sum_{i=1}^m |e_{t+i}|}{\sum_{i=1}^m |Y_{t+i} - Y_{t+i-1}|}. \quad (13)$$

Thiel's U:

$$U = \sqrt{\frac{\sum_{i=1}^m (e_{t+i})^2}{\sum_{i=1}^m (Y_{t+i} - Y_{t+i-1})^2}}. \quad (14)$$

An Overview of Computing these Measures in R [1]

```
pacman::p_load(tidyquant, magrittr, fpp2, xtable)

aapl = tq_get("AAPL", from = "2020-08-17", to = "2020-08-29") %>%
  select(symbol, date, adjusted)
naiveFC = lag(aapl$adjusted)

e = aapl$adjusted - naiveFC
ME = mean(e, na.rm=T)
RMSE = mean(e^2, na.rm=T) %>% sqrt()
MAE = abs(e) %>% mean(na.rm=T)
MPE = 100 * mean(e/aapl$adjusted, na.rm=T)
MAPE= 100 * mean(abs(e)/aapl$adjusted, na.rm=T)
```

An Overview of Computing these Measures in R [2]

```
E = c(ME, RMSE, MAE, MPE, MAPE)
names(E) = c("ME", "RMSE", "MAE", "MPE", "MAPE")
round(E, 2) %>% print()
```

```
##      ME RMSE  MAE  MPE MAPE
## 4.53 9.72 6.98 0.93 1.42
```

Alternatively, we could have just computed it using the fpp2 package

```
accuracy(naiveFC, aapl$adjusted) %>% round(2)
```

```
##              ME RMSE  MAE  MPE MAPE
## Test set 4.53 9.72 6.98 0.93 1.42
```

Outline

- 1 Preface
- 2 Transformations
- 3 Measures of Forecast Accuracy
- 4 Prediction Intervals**
- 5 Recap

Point vs Interval Forecasts

- **Point Forecasts:** future observations for which we report a single forecast observation.
- **Interval Forecast:** a range of values that are reported to forecast an outcome.

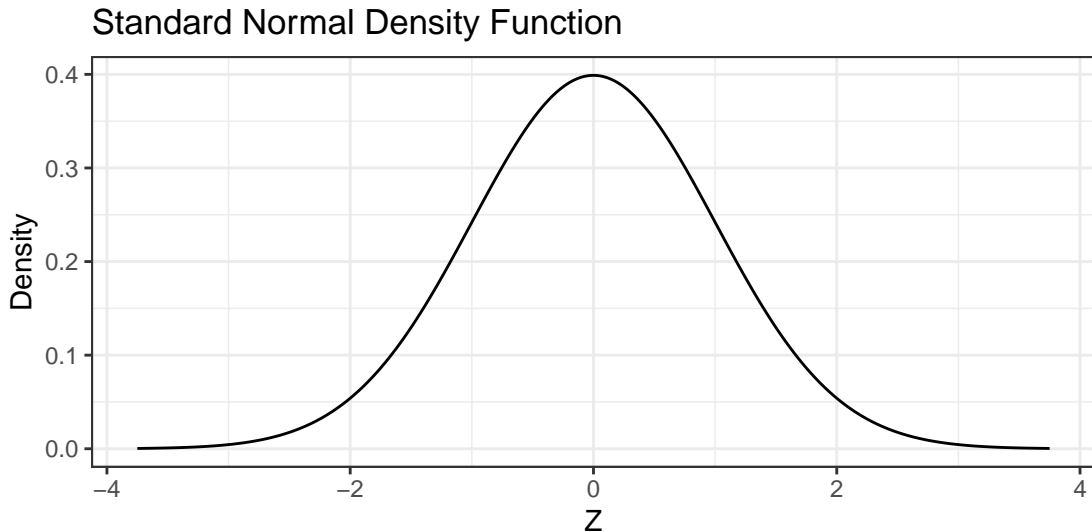
If we assume the forecast errors follow a Normal Distribution, an approximate $100(1 - \alpha)$ prediction interval can be computed as follows:

$$\hat{F}_t \pm Z * RMSE, \quad (15)$$

where:

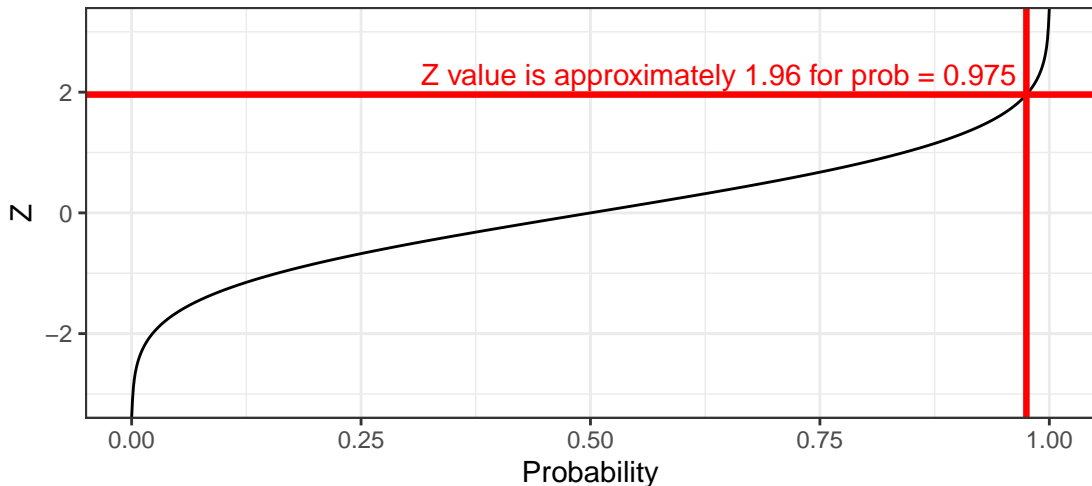
- \hat{F}_t forecast at time t .
- The RMSE can be used as an estimate of the standard deviation of the forecast errors.
- Z is the quantile corresponding to $100(1 - \frac{\alpha}{2})$.

Recall: Standard Normal Distribution [1]



Recall: Standard Normal Distribution [2]

Quantile / inverse CDF of Standard Normal Dist.



Prediction Intervals for the \$AAPL Data [1]

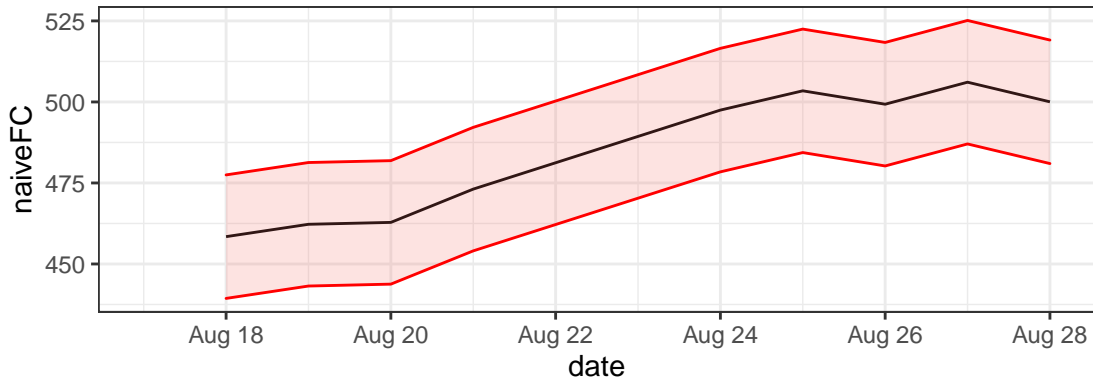
```

PInormU = naiveFC + abs(qnorm(0.975))*RMSE
PInormL = naiveFC - abs(qnorm(0.975))*RMSE
dfNaive = data.frame(date = aapl$date, Ft = naiveFC, PInormU, PInormL)
dfNaive %>% ggplot(aes(x= date, y = naiveFC)) + geom_line() +
  theme_bw() +
  labs(title = "Overlaying 95% Prediction Intervals in ggplot",
       fill = "95% PI") +
  geom_ribbon(aes(ymin = PInormL , ymax = PInormU, fill = "band"),
            alpha = 0.2, color = "red") +
  theme(legend.position= "bottom")

```

Prediction Intervals for the \$AAPL Data [2]

Overlaying 95% Prediction Intervals in ggplot



95% PI  band

Outline

- 1 Preface
- 2 Transformations
- 3 Measures of Forecast Accuracy
- 4 Prediction Intervals
- 5 Recap**

Summary of Main Points

Main Learning Outcomes

- Apply (log) transformations to a time series.
- Apply and interpret measures of forecast accuracy.
- Interpret prediction intervals for a simple forecast.

Things to Do

- Thoroughly read all of Chapter 2 of our book.
- Go through the slides, examples and make sure you have a good understanding of what we have covered.
- Recreate the plots in Slides 15-17 for the [RSCCASN Data](#). Note that the period, $m = 12$.
- If you are interested in additional practice problems, please consider the following problems from your textbook. To access these datasets, please click [here](#).
 - For the Means approaches in Example 2.7 (P.49), use R to compute the 7 error measures for the four forecasting approaches.
 - Exercise 2.12 and compute the forecast errors for the naive forecast.

Graded Assignment 04: Evaluating your Retention/Focus

Please go to [Canvas \(click here\)](#) and answer the questions. **Due Sept. 3, 2020 [2:50 PM, Ohio local time]**.

What/Why/Prep? The purpose of this assignment is to evaluate your understanding and retention of the material covered up to the end of Class 04. In order to prepare for this, you should have either actively attended class and/or watched the recording from WebEx. Furthermore, you should have thoroughly read up to the end of Chapter 2 from your textbook.

General Guidelines:

- Individual assignment.
- This is **NOT** a timed assignment (i.e. once you start the assignment you will have 10-15 minutes to complete the one question).
- Proctorio is NOT required for this assignment.
- You will need to have R installed (or accessible through the [Remote Desktop](#))

Graded Assignment 05: Evaluating your Retention/Focus

Please go to [Canvas \(click here\)](#) and answer the two questions. **Due Sept. 7, 2020 [2:50 PM, Ohio local time]**.

What/Why/Prep? The purpose of this assignment is to evaluate your understanding and retention of the material covered up to the end of Class 05. In order to prepare for this, you should have either actively attended class and/or watched the recording from WebEx. Furthermore, you should have thoroughly read up to the end of Chapter 2 from your textbook.

General Guidelines:

- Individual assignment.
- This is **NOT** a timed assignment (i.e. once you start the assignment you will have 10-15 minutes to complete the one question).
- Proctorio is NOT required for this assignment.
- You will need to have R installed (or accessible through the [Remote Desktop](#))

ISA 444: Business Forecasting

05 - Basic Tools and Goodness of Fit (Cont.)

Fadel M. Megahed

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University
Email: fmegahed@miamioh.edu
Office Hours: [Click here to schedule an appointment](#)

Fall 2020