Math 361

Day 19

Project, Methods Table and

Inv. 2.1-Quantitative data

Announcements

- Group Project due May 22, revised by June 5

Learning Objectives

1. State the purpose of a **descriptive** and an **inferential** statistical method

2. Describe the *center, shape, and spread* of a single quantitative variable from a **dotplot or histogram**

3. Interpret the values of the **mean** and **standard deviation** computed from a single quantitative variable

Recall types of variables

- Binary
- Categorical
- Numerical/quantitative

In Chapter 1, we looked at **one binary variable**

In Chapter 2, we'll look at **one quantitative variable**

Descriptive vs. Inferential Methods

Descriptive methods are used to describe what is in a dataset

Inferential methods are used to infer a population characteristic

For a **single binary variable**, we've used a

- Bargraph
- Sample proportion $\hat{p}$
- Binomial Test
- 1-sample z-test for proportions
- Wald 95% confidence interval
- Plus Four confidence interval

Classify each as **descriptive** or **inferential**

Descriptive vs. Inferential Methods

Descriptive methods are used to describe what is in a dataset

Inferential methods are used to infer a population characteristic

For a single binary variable, we've used a

- Bargraph
- Sample proportion $\hat{p}$
- Binomial Test
- 1-sample z-test for proportions
- Wald 95% confidence interval
- Plus Four confidence interval

Classify each as **descriptive** or **inferential**

Descriptive Methods for One Quantitative Variable

Recall from Inv. A that the ***center, spread and shape*** of the distribution of the variable are of interest.

Shape **via Graphs:** dotplot, histogram

Numerical Summaries* of *center*: mean, median, mode

Numerical Summaries* of *spread*: standard deviation, variance, range, IQR

* Remember a "statistic" is a numerical summary of a sample

Investigation 2.1 (p. 134)

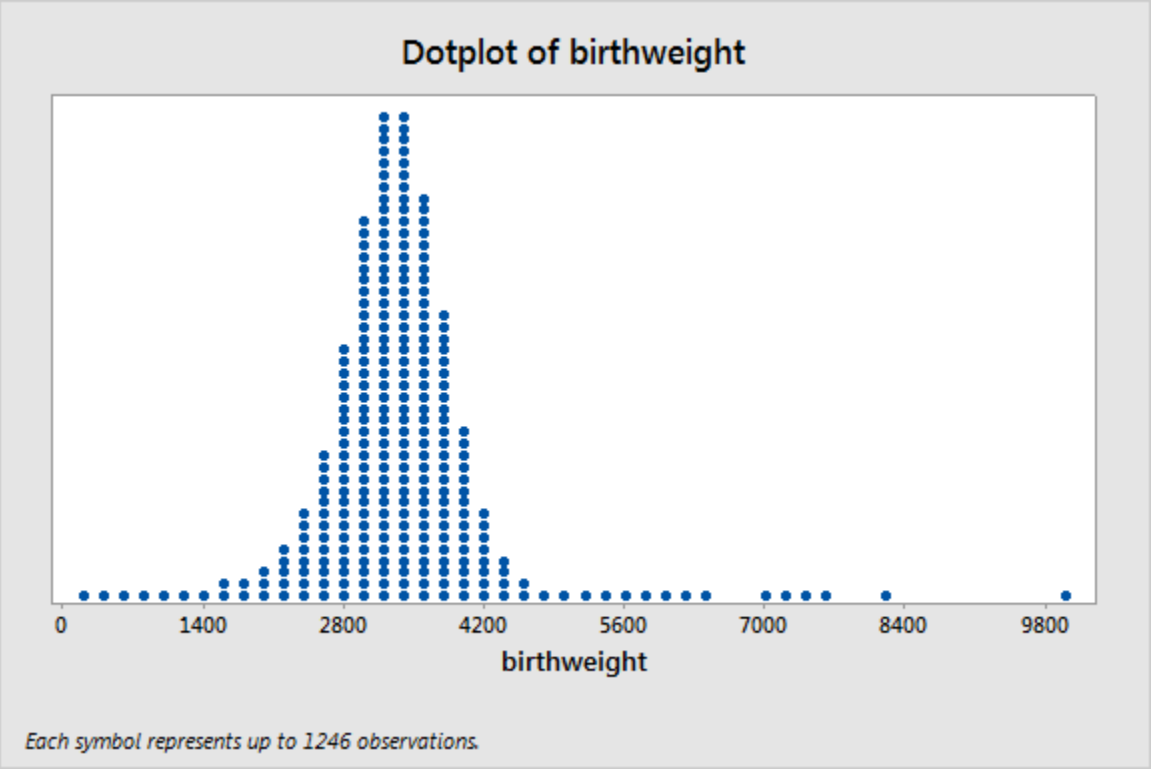
Use “USBirthsJan2013.txt” from daily schedule:

It’s too big for our applet so let’s switch to Minitab or R

n=324,314 Babies

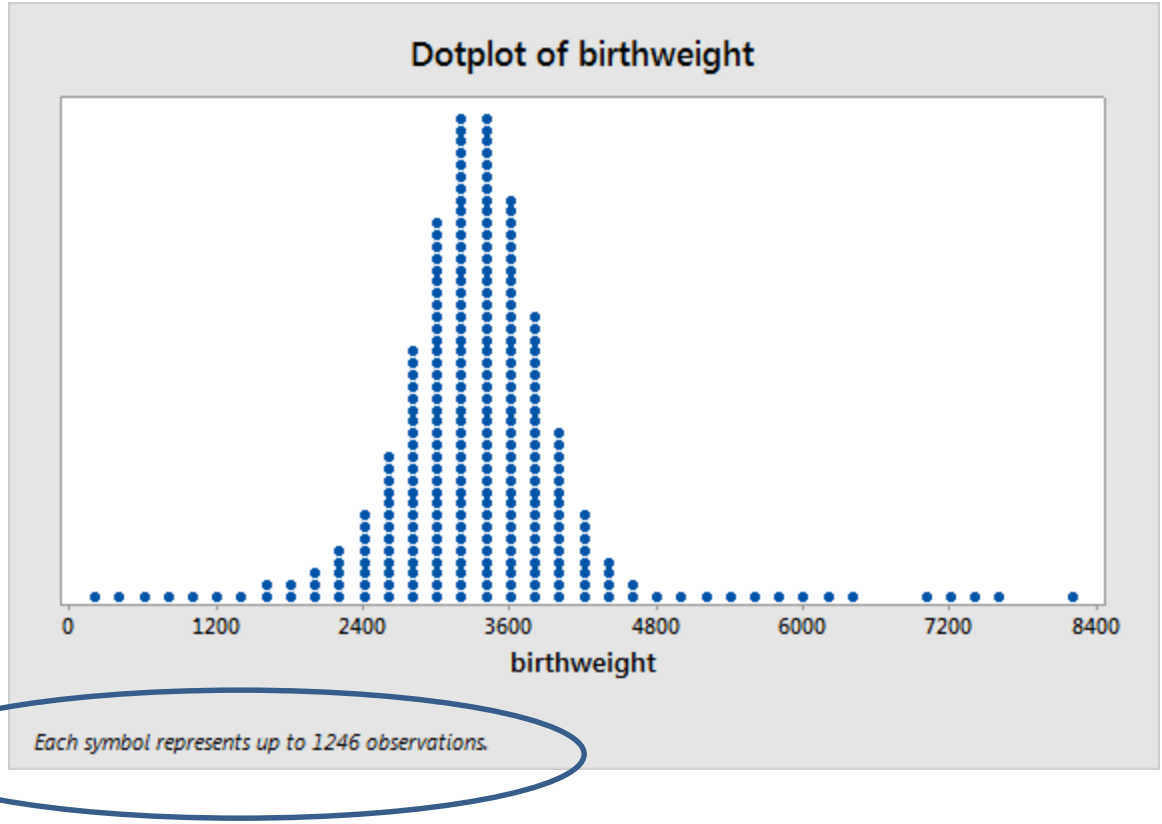
Part c)

what do you see?

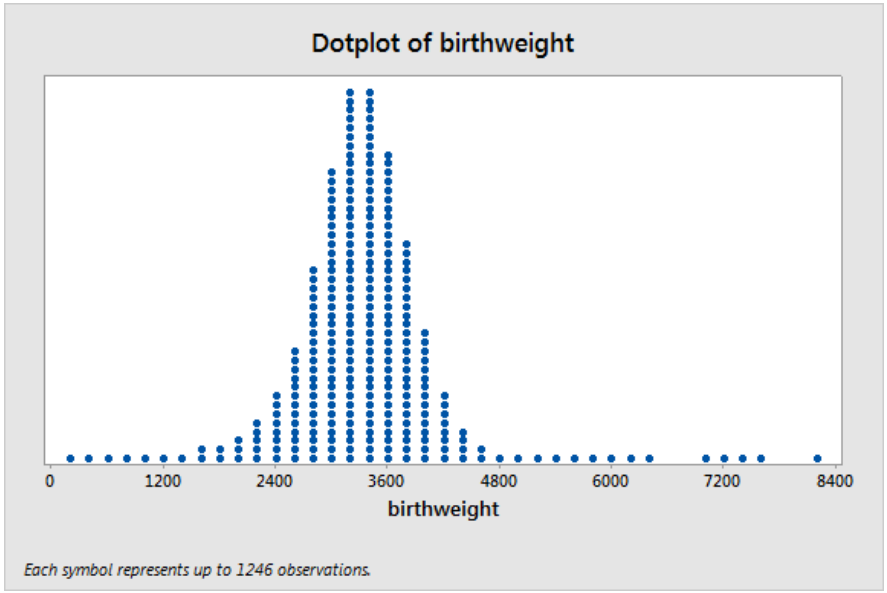


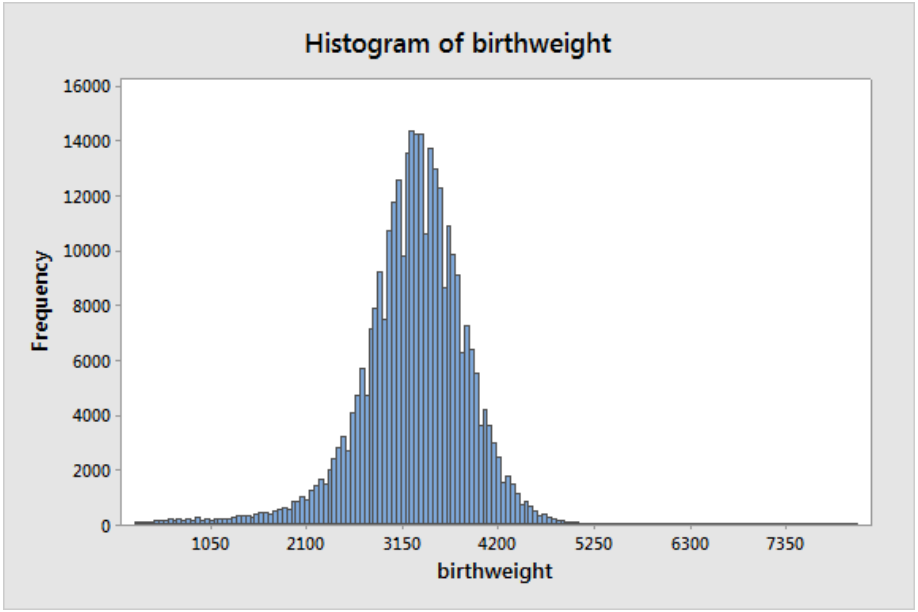
Inv. 2.1, part (d)

Removed “9999” values – the code for “missing”



Inv. 2.1, part e: Dotplot vs. Histogram





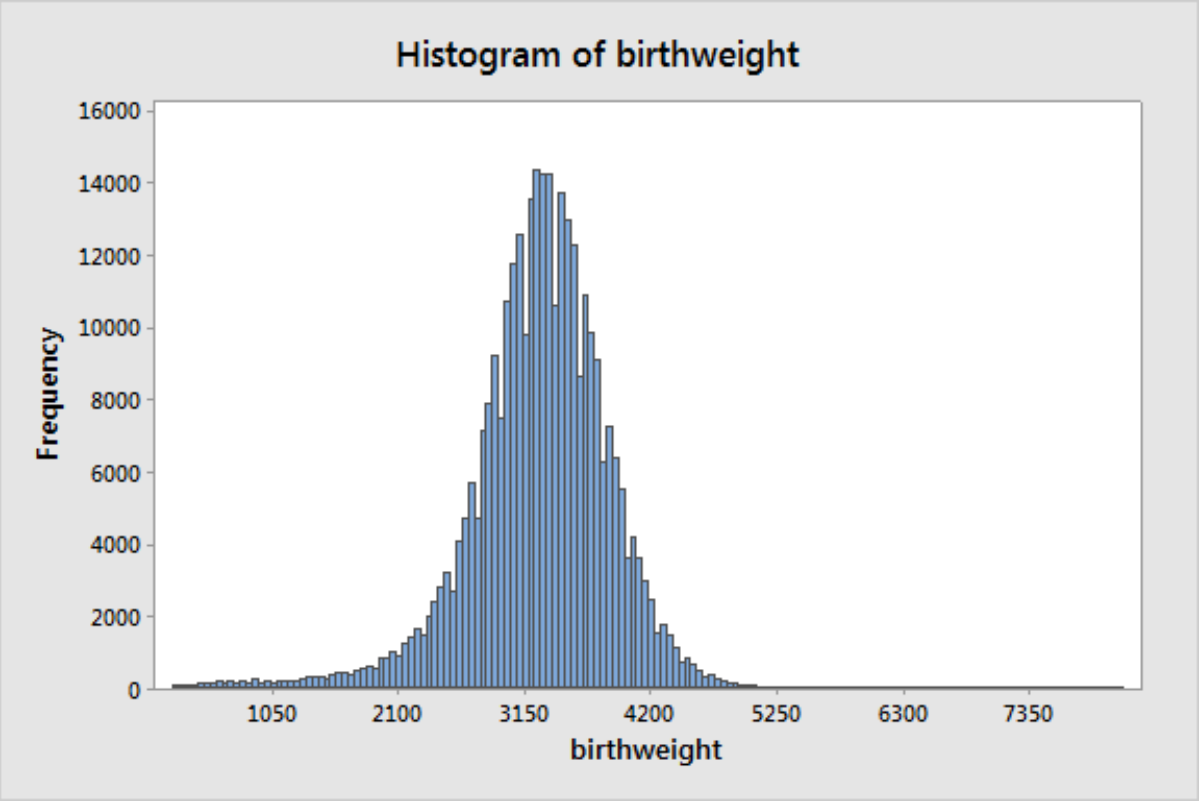
Note: Minitab has automatically “binned” the dotplot so that the information revealed by the histogram and dotplot is similar: other softwares may not bin dotplots.

Inv. 2.1, part (f): describe the distribution

Center?

Spread?

Shape?



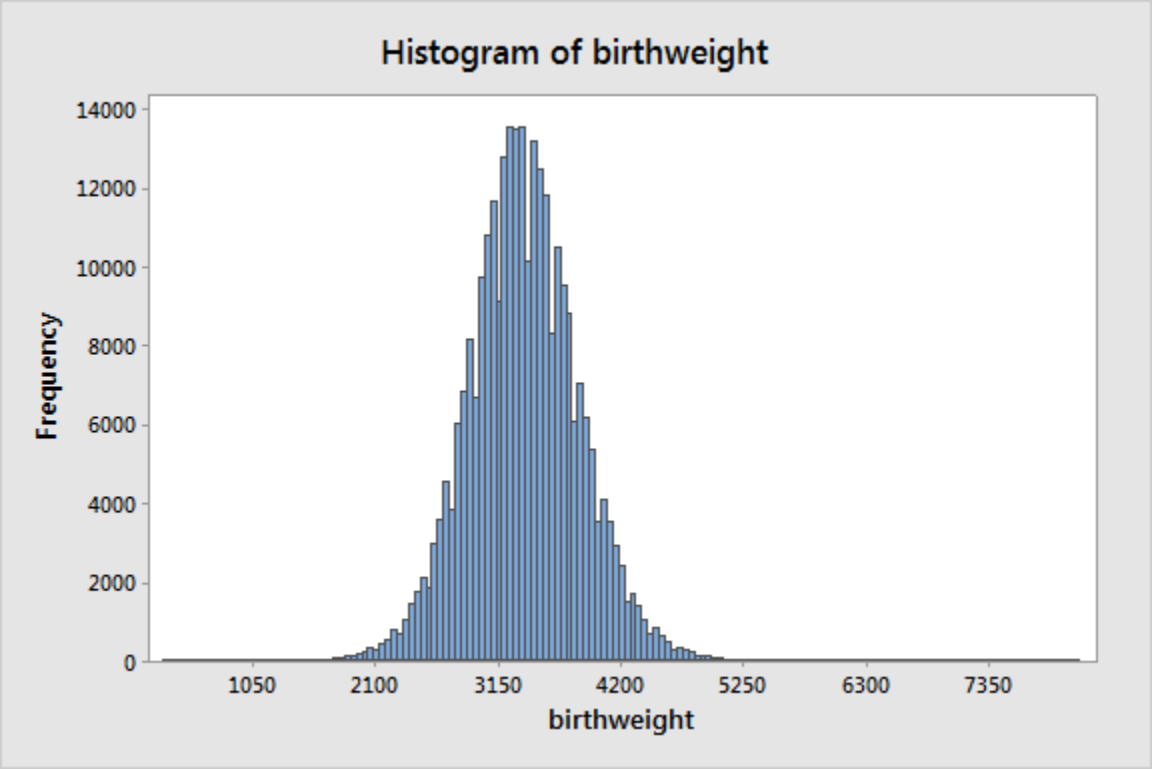
Inv. 2.1, part (h) & (i): Full Term Babies

n=285,907

Center?

Spread?

Shape?



Inv. 2.1 part (j)

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
birthweight	285907	0	3361.6	0.888	474.6	312.0	3060.0	3350.0	3657.0	8115.0

Interpret the **mean** and **standard deviation**:

Inv. 2.1 part (k)

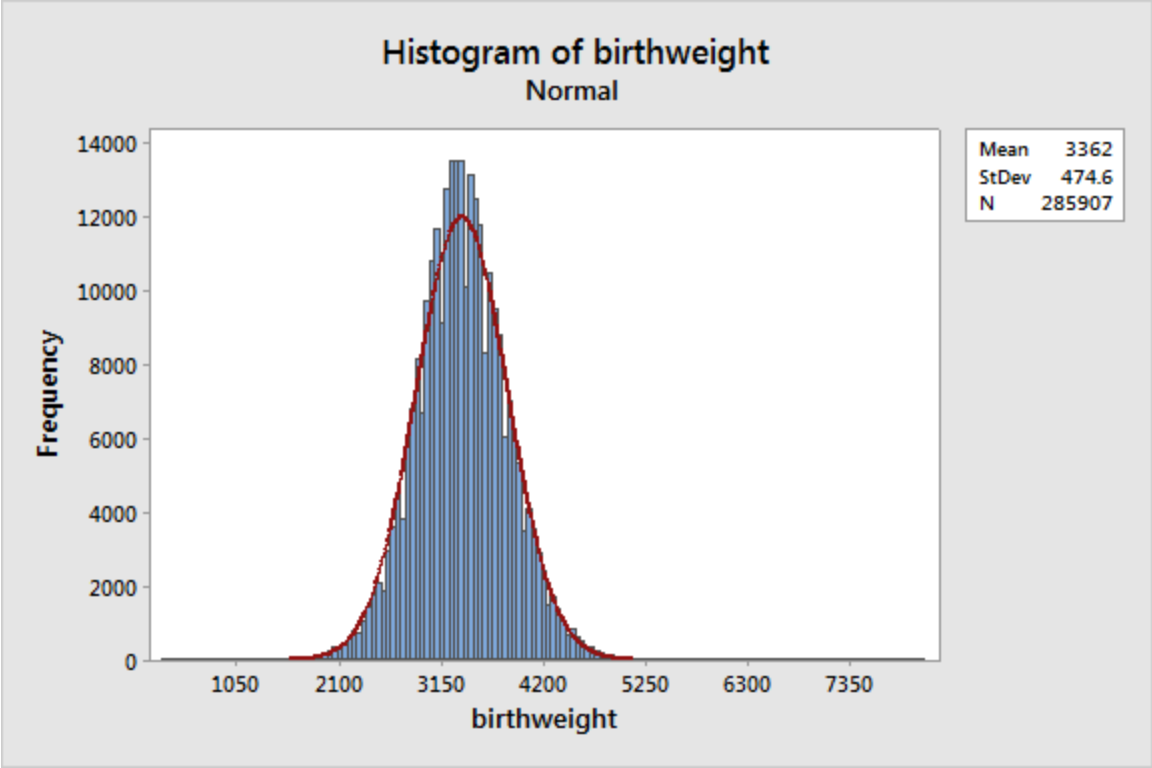
Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
birthweight	285907	0	3361.6	0.888	474.6	312.0	3060.0	3350.0	3657.0	8115.0

How would the **mean** and **standard deviation** change if we put back the 9999 values? The pre-term babies?

Inv. 2.1 part (m)

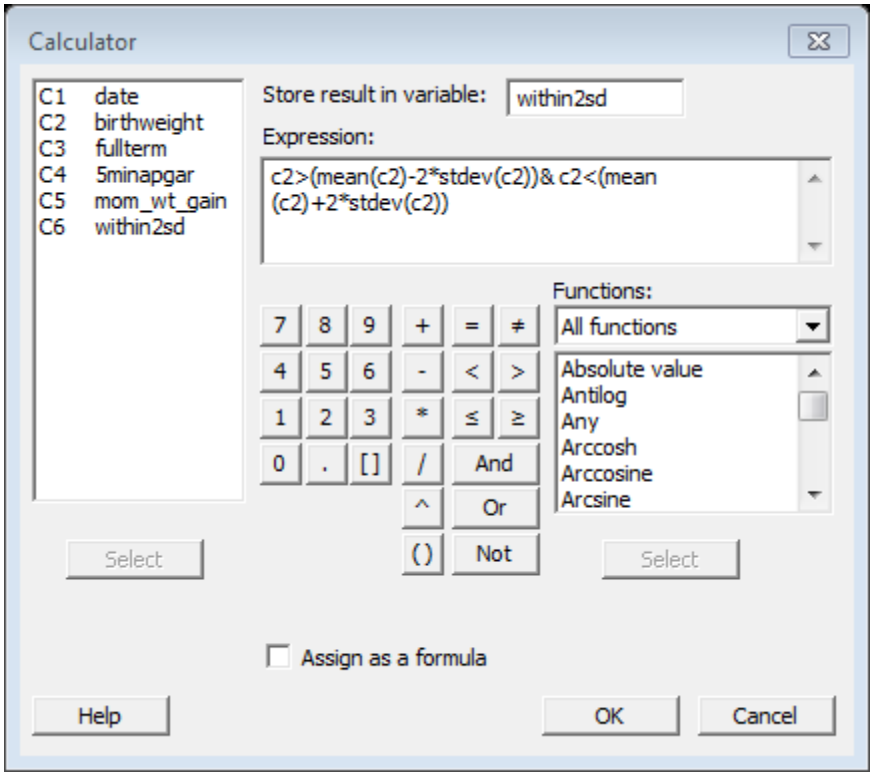
Remember the magical **NORMAL DISTRIBUTION**?



Inv. 2.1 part (n)

Recall that the Empirical Rule states that about 95% of observations in a Normal Distribution are within 2 SDs of the mean.

Let's compute the proportion of Birthweights that are within 2 SDs of the mean…



Inv. 2.1 part (n)

What proportion of birth weights are within 2 SDs of the mean?

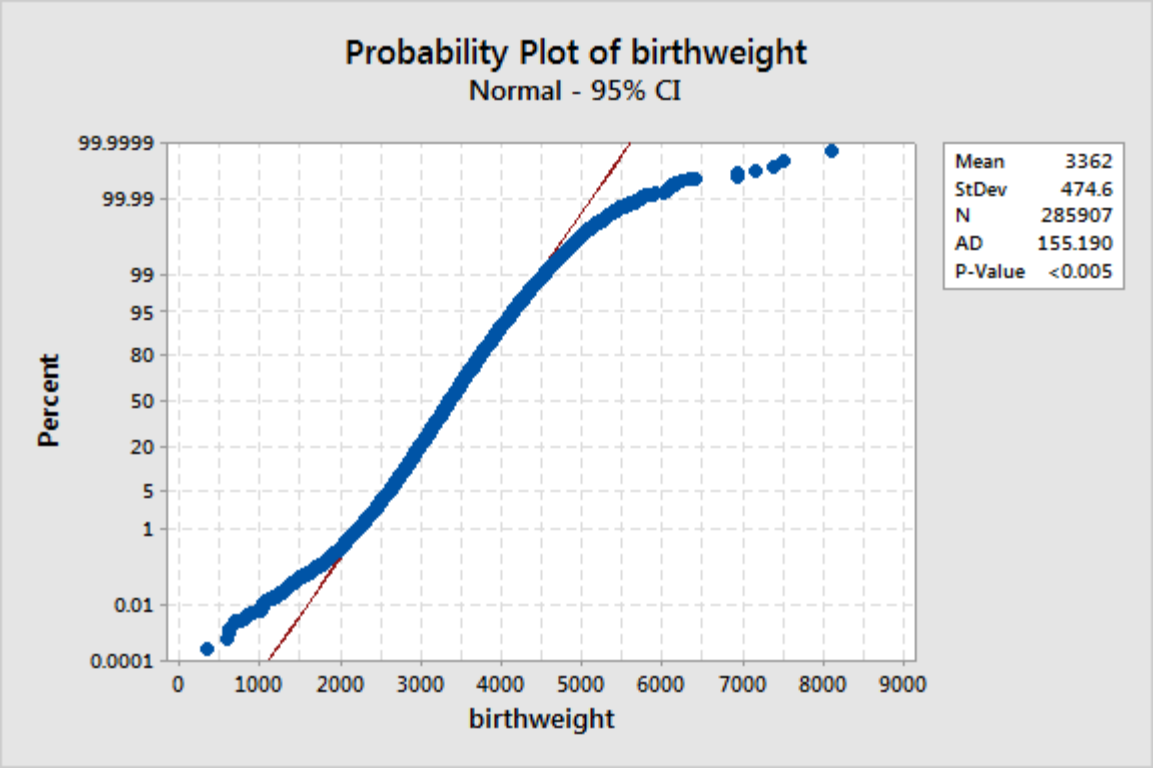
Tally for Discrete Variables: within2sd

Tally

within2sd	Count	Percent
0	13720	4.80
1	272187	95.20
N=	285907	

Inv. 2.1 part (p)

Normal Probability Plot – can be used to assess normality



Summary of Inv. 2.1

Descriptive Statistics for 1 quantitative variable