Chapter 5: The standard deviation as a ruler and the normal model p131

Which is the better exam score?
- – 67 on an exam with mean 50 and SD 10
- – 62 on an exam with mean 40 and SD 12?

Is it fair to say:
- – 67 is better because 67 > 62?

- – 62 is better because it is 22 marks above the mean and 67 is only 17 marks above the mean?

Key: *z-scores*.

Effect of a Linear Transformation on summary statistics

•Multiplying each observation in a data set by a number *b* multiplies the mean, median, by *b* and the measures of spread (standard deviation, *IQR*) by *abs(b)* .

•Adding the same number *a* to each observation in a data set adds *a* to measures of center, quartiles, percentiles but does not change the measures of spread.

Standardizing and *z*-scores p132

    If *x* is an observation from a distribution that has mean μ and std. dev. σ , the standardized value of *x* is

$$z = \frac{x - \mu}{\sigma}$$ A standardized value is often called a *z*-score.

    A *z*-score tells us how many standard deviations the original observation falls away from the mean.

What are the mean and SD **of z**?

No matter what mean and SD x has, z has mean 0, SD 1.

- Calculating a z-score sometimes called "standardizing". Above says why.
- Gives a basis for comparison for things with different means and sds.

Those exam scores above:

Which is the better exam score?
- 67 on an exam with mean 50 and SD 10
- 62 on an exam with mean 40 and SD 12?

Turn them into z-scores:
- 67 becomes (67-50)/10=1.70
- 62 becomes (62-40)/12=1.83

so the 62 is a (slightly) better performance, relative to the mean and SD.

Density Curve and the Normal Model (p137)

Density curve is a curve that

-   is always on or above the
    horizontal axis.

- has area exactly 1 underneath it.

•A density curve describes the overall pattern of a distribution.

Example: The curve below shows the density curve for scores in an exam and the area of the shaded region is the proportion of students who scored between 60 and 80.
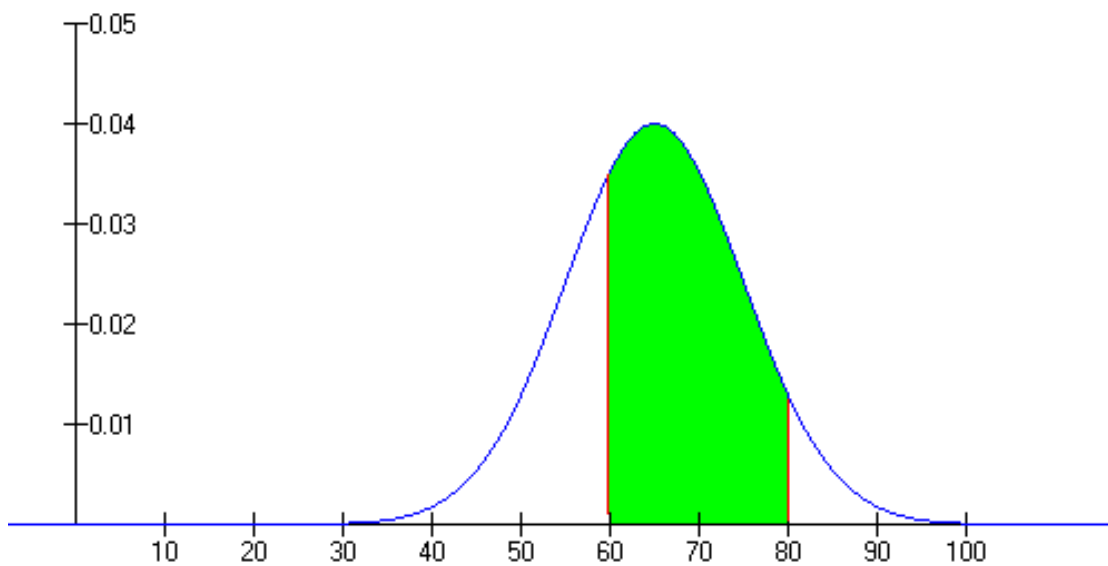
Normal distributions p139

•An important class of density curves are the symmetric unimodal bell-shaped curves known as **normal curves**. They describe **normal distributions**.

•All normal distributions have the same overall shape.

•The density curve for a particular normal distribution is specified by giving the mean μ and the standard deviation σ.

•The mean is located at the center of the symmetric curve and is the same as the median (and mode).

The standard deviation σ controls the spread of a normal curve.

• There are other symmetric bell- shaped density curves that are not normal.

• The normal density curves are specified by a particular equation. The height of a normal density curve at any point x is given by

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

According to data released by chamber of commerce of a city, the weekly wages of factory workers are Normaly distributed with mean $700 and standard deviation $60.

What proportion of the factory workers in this city makes a weekly wage less than $600?

Ans Z = (600-700)/60 = -1.67

look up -1.67 in table Z to get 0.0475.

What proportion of the factory workers in this city makes a weekly wage less than $750?

And z=(750-700)/60=0.83

0.7867

What proportion of the factory workers in this city makes a weekly wage greater than $850?

And z=(850-700)/60=2.50

0.9938

1-0.9938 = 0.0062

What proportion of the factory workers in this city makes a weekly wage greater than $600 and $750?

0.7867 - 0.0475 = 0.7392

Getting values from proportions

- Use Table Z backwards to get z that goes with proportion less
- Turn z back into original scale by using x=mean + SD * z.
- Why? z=(x-mean)/SD, solve for x

The marks in an examination have a distributed with a mean of 55 marks and a standard deviation of 10 marks.

A student just realized that only the bottom 5% of candidates failed the examination?

What is the minimum grade required to pass this exam?

For the 5%, z = 1.645 and so x = 55 - 1.645 x 10 = 38.55.

If the top 20% of candidates obtain qualify for an award, what is the minimum grade required in order to quality for this award?

Z value for 0.80 = 0.84 and so x = 55+0.84*10 = 63.5

Normal quantile plots p148

•A histogram or stem plot can reveal distinctly nonnormal features of a distribution.

•If the stemplot or histogram appears roughly symmetric and unimodal, we use another graph, the normal quantile plot as a better way of judging the adequacy of a normal model
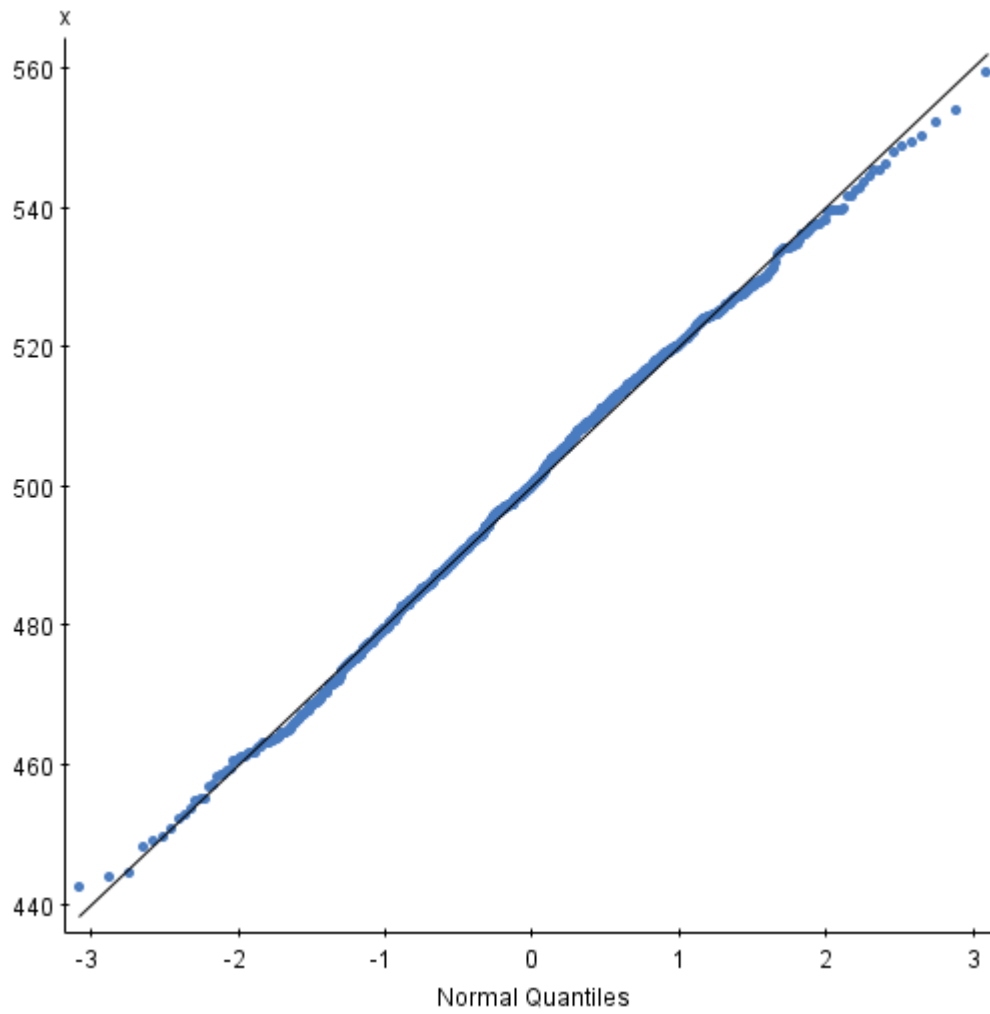
•Use of normal quantile plots.

If the points on a normal lie close to a straight line, the plot indicates that the data are normal.

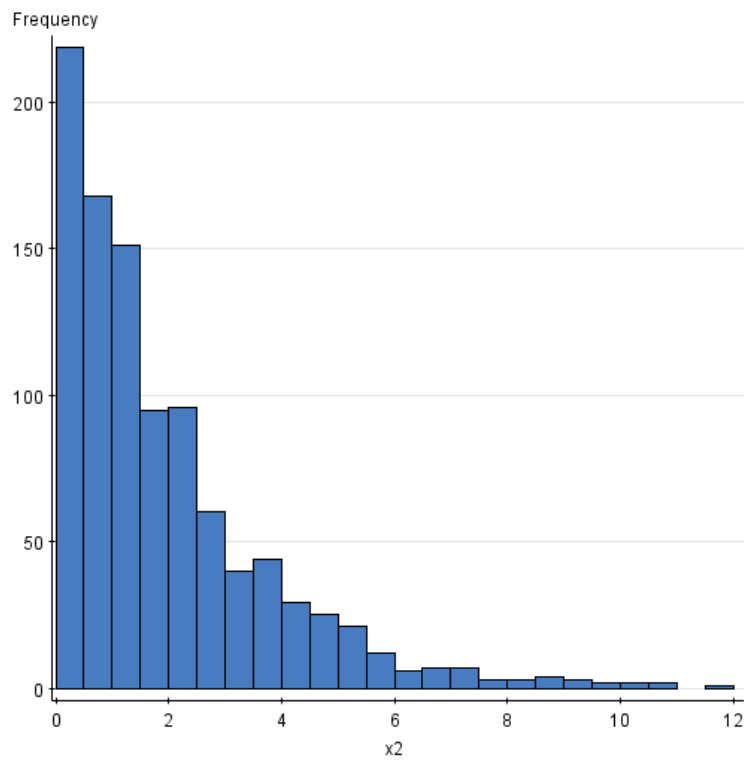Outliers appear as points that are far away from the overall pattern of the plot.

Histogram and the nscores plot for data generated from a normal distribution (N(500, 20)) (for 1000 observations)
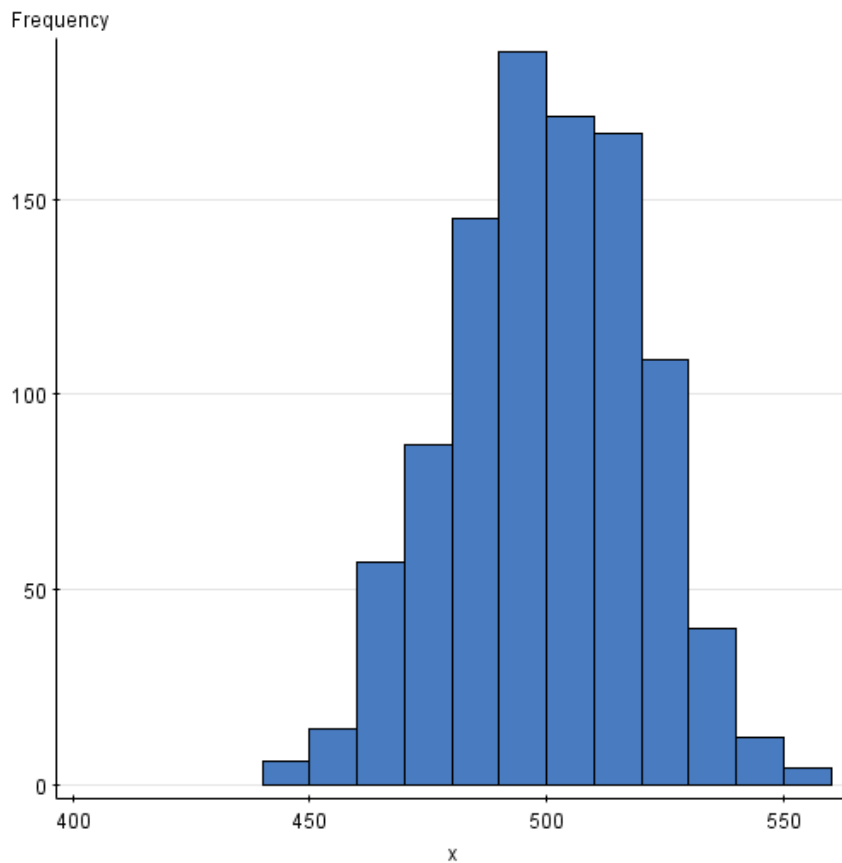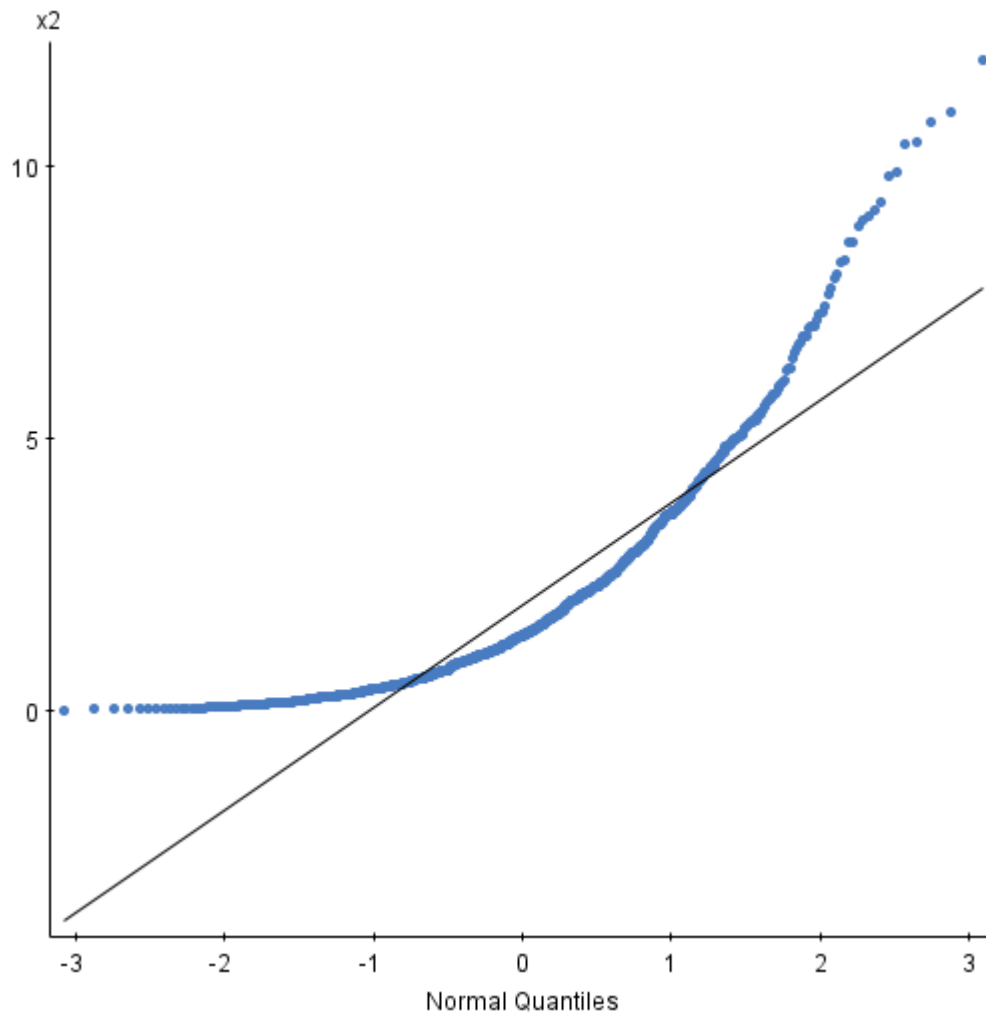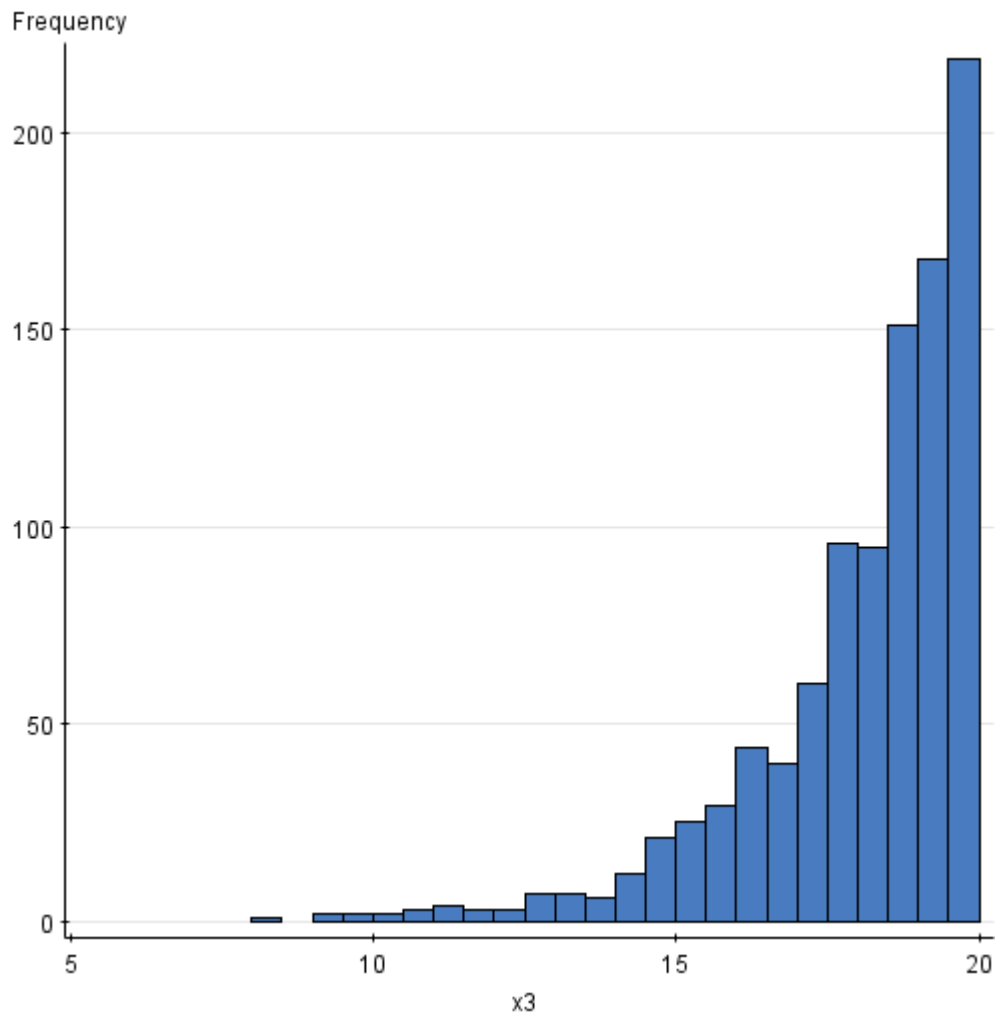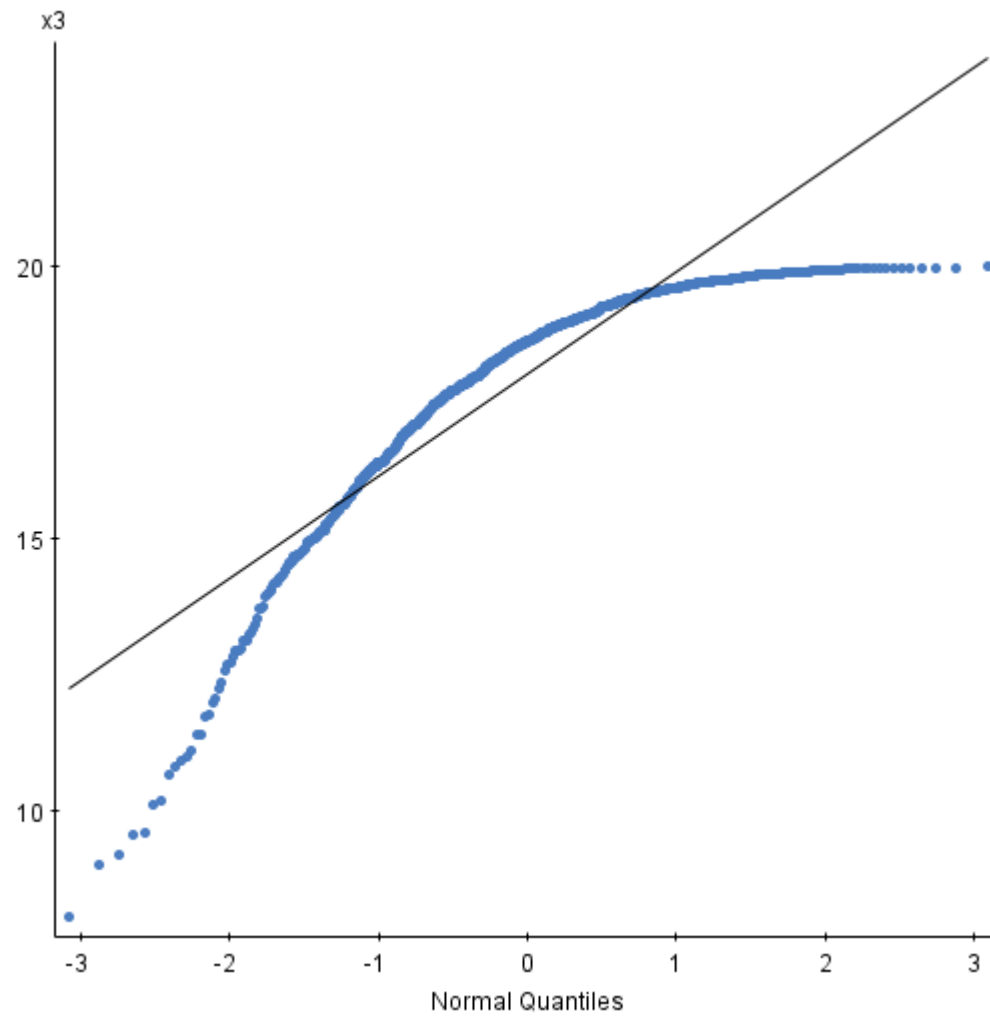
Histogram and the nscores plot for data generated from a right skewed distribution

Histogram and the nscores plot for data generated from a left skewed distribution

•**The 68-95-99.7 rule p141**

In the normal distribution with mean m and std. deviation s,

■ Approx. 68% of the observations fall within one standard deviation of the mean.

■Approx. 95% of the observations fall within two standard deviations of the mean .

■Approx. 99.7% of the observations fall within 3 standard deviations of the mean.

•Example

The distribution of heights of women aged 18-24 is approx. normal with mean μ = 64.5 inches and std. dev.

σ = 2.5 inches.

2σ = 5 inches. The 68-95-99.7 rule says that the middle 95% (approx.) of women are between 64.5-5 to 64.5+5 inches tall.

The other 5% have heights outside the range from 64.5-5 to 64.5+5 inches .

2.5% of the women are taller than 64.5+5 .

Ex. 1) The middle 68% (approx.) of women are between _____ to ____ inches tall.

2) ____% of the women are taller than 67

3) ___% of the women are taller than 72