Modeling with IID Normal Draws

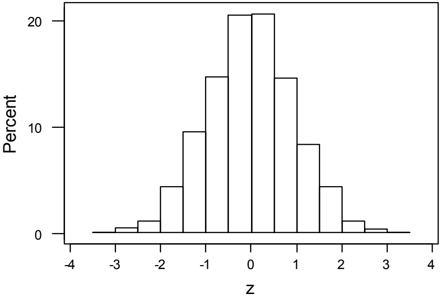
Rob McCulloch

1. The Histogram and IID Draws

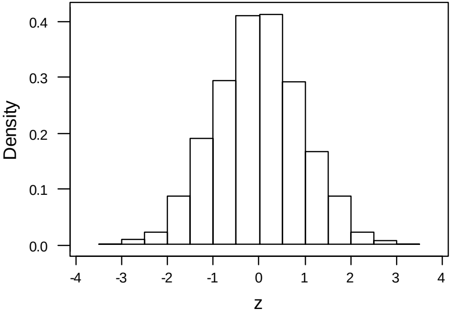
Here is the histogram of 1000 draws from the standard normal.

The height of each bar tells us the number of observations in the interval.

All the intervals have the same width.



If we divide
the height of
each bar
by the width*1000
the picture looks
the same, but
now the area of
each bar =
% of observations
in the interval
(trust me).



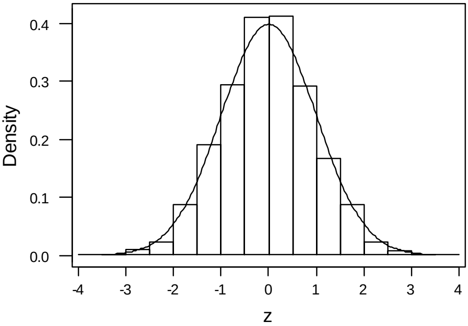
It **looks the same**, but the vertical scale is different.

For a large number of iid draws, the observed percent in an interval should be close to the probability.

For the density the area is the probability of the interval.

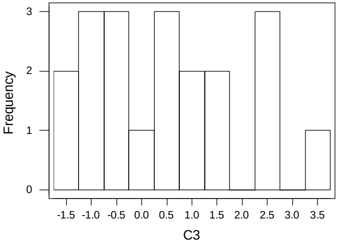
For the histogram the area is the observed percent in the interval.

In large samples these are close.

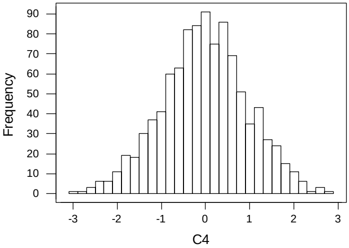


Example

histogram of 20
i.i.d N(0,1) draws



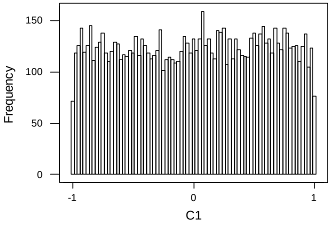
histogram of 1000
i.i.d N(0,1) draws



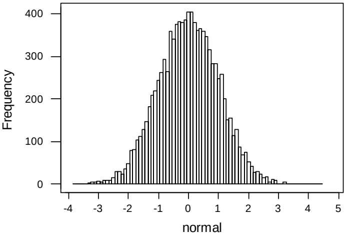
The histogram of a “large” number of i.i.d draws from any distribution should look like the p.d.f.

Example

10,000 draws, uniform on (-1,1)

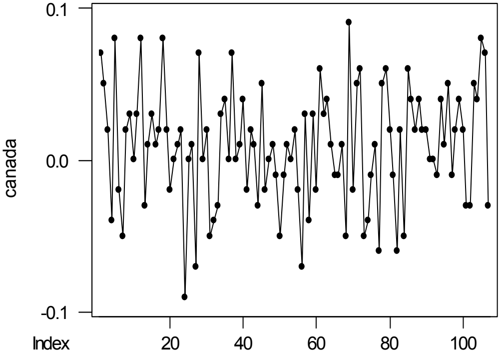


10,000 draws, N(0,1)

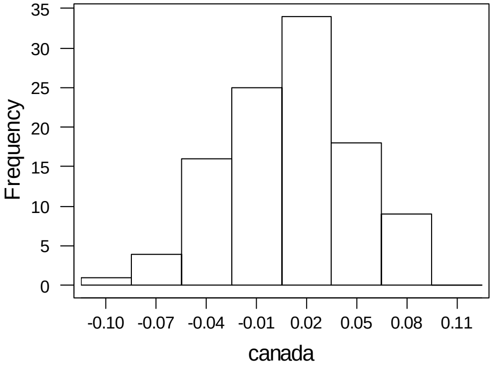


2. The Normal Distribution and Data

We look at the return data for Canada.
We have monthly returns from Feb ‘88 to Dec ‘96.



No apparent pattern!



Normality seems reasonable!

Conclude: The returns look like i.i.d normal draws!

Estimation of Parameters

Our model is

$$R_t \sim N(\mu, \sigma^2), \text{IID}$$

We have two parameters we have to estimate, μ and σ.

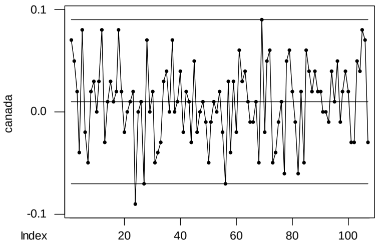
We will estimate the true long run mean with the sample mean $\hat{\mu} = .01$.

We will estimate the true long run standard deviation with the sample standard deviation $\hat{\sigma} = .04$.

If we think μ is about .01 and σ is about .04 (based on the data), then our best guess at the next return is .01.

An interval which has a 95% chance of containing the next return would be:

$$.01 +/- .08$$



Our ***model*** is

$R_t \sim N(.01,.04^2)$ iid

Note: we used i.i.d. Bernoulli draws to model coin tosses and defects.

Now we are using the idea of i.i.d. normal draws to model returns! We have a statistical model for the real world.

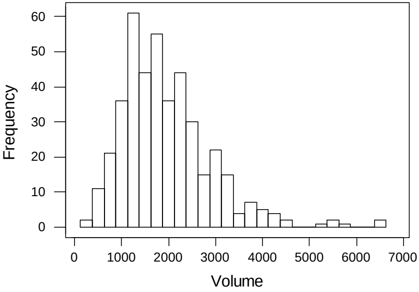
This is a powerful statement about the real world.

Example

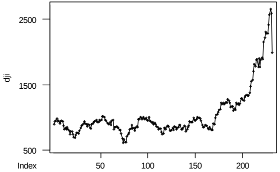
Of course, not all data looks normal.

Daily volume
of trades
in the Cattle
pit.

Skewed to the right.

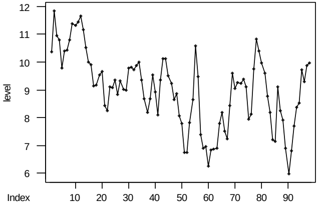


Dow Jones

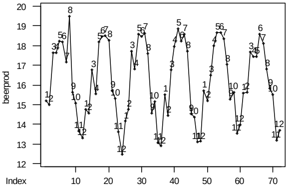


and not all time series look independent:

Lake Level



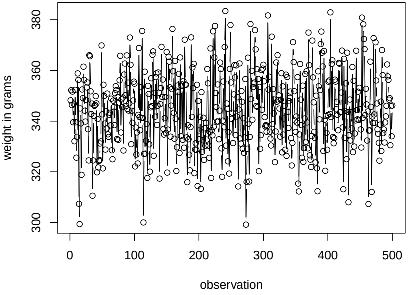
Beer Production

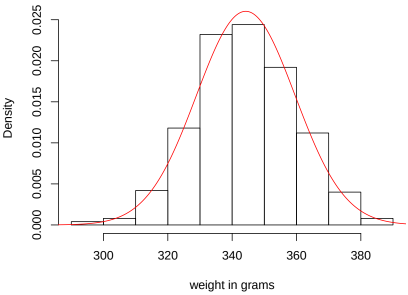


The Cereal Data:

Your job is to run the process that fills cereal boxes with cereal.

You can't get it to put the same amount of cereal in each box!!

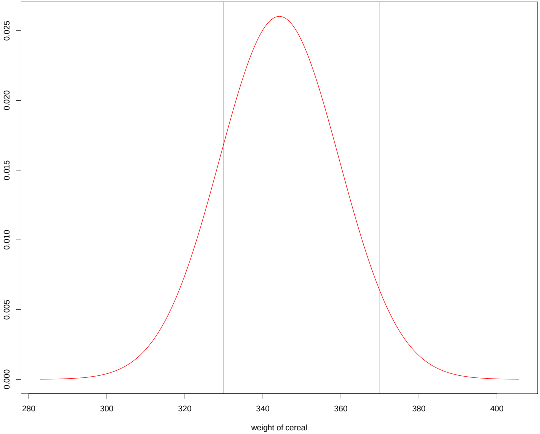




The time-series plot and the histogram show the amount of cereal (weight in grams) put into 500 boxes.

They look like iid Normal draws !!

We chose μ to be the sample mean which is 344.22 and σ to be the sample standard deviation which is 15.33.



You are told an "acceptable" amount of cereal is 350 ± 20.

How are you doing?

What is the probability that the next box has an acceptable amount of cereal?

CDF at 370 - CDF at 330 = .9537 - .177 = .78.

Prob next one is bad is .22.

What is the prob next 50 are good?

$.78^{50} = 4.024669e - 06$.

Not too good!

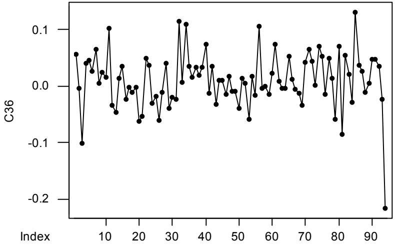
3. Standardization

How unusual is it?

Sometimes something weird or unusual happens and we want to quantify just how weird it is.

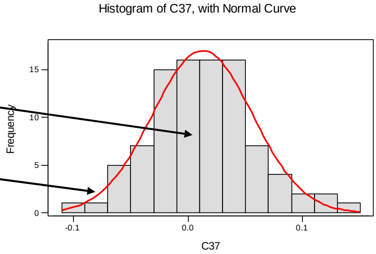
A typical example is a market crash.

Monthly returns on a market index from Jan ’80 to Oct ’87.



Question: how crazy is the crash?

The data looks normal (from the histogram).
We can use a normal curve (model) to describe all the values ***except the last.***

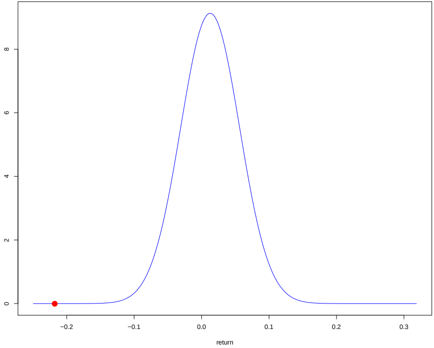


The curve (model) has μ=.0127 and σ=.0437.

We are "estimating" the true μ and σ based on the data. We will be very precise in the future.

The crash month return was -.2176.

Here is the pdf of $N(.0127, .0437^2)$ with red dot at the crash return.

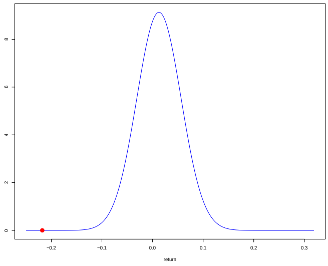


Wow, the return is way out in the left tail!

We would like to quantify how unusual the crash is given the normal distribution.

Sometimes people report the chance of getting something that far out.

In this case, we could evaluate the $N(.0127, .0437^2)$ cdf at -.2176 which is 5.610441e-07 (*tiny*).



Another thing people report is the *z-value* which is the how many standard deviations away from the mean the value is:

For $X \sim N(\mu, \sigma^2)$, the **z value** corresponding to an observed x value is

$$z = \frac{x - \mu}{\sigma}$$

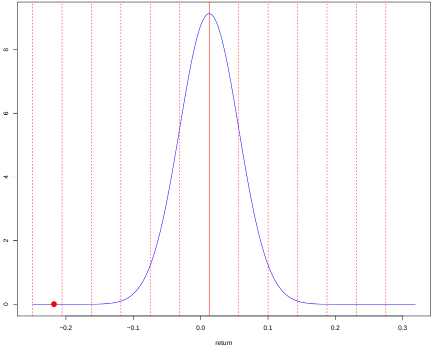
The z value tells us how many standard deviations away from the mean the observed x value is.

For our crash example

$$z = \frac{x-\mu}{\sigma} = \frac{-.2176 - .0127}{.0437} = -5.27.$$

Center line at μ.

other lines at
$\mu \pm i\,\sigma,\ \ i = 1, 2..6.$



Wow! the return for the crash month was over 5 standard deviations away from the mean !!!!

Standardization:

Another way to look at it is that the z value is a draw from the standard normal:

$$X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

If X should look like a normal draw, then Z should look like a standard normal draw.

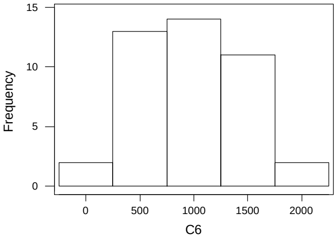
crash example:
getting -.2176 from $N(.0127, .0437^2)$ is *just like* getting -5.27 from a standard normal.

Example

How unusual is (the hockey player) Wayne Gretzky ?
(Recall that ESPN picked him 5^{th} greatest athlete of the century).

This is the histogram of the total career points of the 42 players judged by the Hockey News to be the greatest ever, not counting goalies and Wayne Gretzky.

μ=1000 and σ = 450 look reasonable (based on the data).



Gretzky had 2855 points.

```
MTB > let k1 = (2855-1000)/450

MTB > print k1
```

Data Display

```
K1    4.12222
```

Grezky is like getting 4.1 from the standard Normal. No way!