

Uncertainties01.wmxm

Table of Contents

Preface	1
References	1
fracData (List, a, b)	3
confidence (q, m, s)	3
Basic Concepts in Error Analysis	3
Spring-loaded Projectile Launcher Experiment	4
Precision and Accuracy Defined	8
20% of measurements incorrectly logged example	9
Rules of Thumb for Precision	12
Achieving Accurate Measurements	13
Measurement Errors and the Central Limit Theorem	13
Why Errors are Gaussian	14
Expectation $E(F)$ and Variance $V(F)$ Defined	14
Central Limit Theorem (CLT)	17
Justification of the Central Limit Theorem	18

1 *Preface*

In Uncertainties01.wmxm we introduce basic concepts in error analysis, with an emphasis on the physical sciences.

Edwin L. (Ted) Woollett
<https://home.csulb.edu/~woollett/>
 June 3, 2024

2 *References*

[HH] Ifan Hughes, Thomas Hase, *Uncertainties: A Practical guide to Modern Error Analysis*, 2010, Oxford Univ Press.

R. J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences (Manchester Physics Series)*, 1993, Wiley

Gerhard Bohm, Günter Zech,
Introduction to Statistics and Data Analysis for Physicists, 3rd revised edition, 2017,
 available from CERN webpage:
<https://s3.cern.ch/inspire-prod-files-d/da9d786a06bf64d703e5c6665929ca01>

Frederick James, *'Statistical Methods in Experimental Physics'*, 2nd ed.,
 2006, World Scientific.

Louis Lyons, *Statistics for Nuclear and Particle Physics*, 1986, Cambridge Univ. Press,

Luca Lista, *'Statistical Methods for Data Analysis in Particle Physics'*,
Lecture Notes in Physics 909, 2016, Springer-Verlag,

Glen Cowan , *Statistical Data Analysis*, Clarendon Press, Oxford, 1998

M.I.T. Course: *Fundamentals of Statistics*, (formerly *'Statistics for Applications'*),
 Philippe Rigollet lectures on YouTube

1. https://www.youtube.com/watch?v=VPZD_aj8H0
2. https://www.youtube.com/watch?v=C_W1adH-NVE

List of YouTube videos:

https://www.youtube.com/playlist?list=PLUI4u3cNGP60uVBMaoNERc6knT_MgPKS0

<https://ocw.mit.edu/courses/18-650-statistics-for-applications-fall-2016/>

lecture slides:

[https://ocw.mit.edu/courses/18-650-statistics-for-applications-fall-2016/
 pages/lecture-slides/](https://ocw.mit.edu/courses/18-650-statistics-for-applications-fall-2016/pages/lecture-slides/)

lecture videos:

[https://ocw.mit.edu/courses/18-650-statistics-for-applications-fall-2016/
 video_galleries/lecture-videos/](https://ocw.mit.edu/courses/18-650-statistics-for-applications-fall-2016/video_galleries/lecture-videos/)

lecture assignments (pdf's):

[https://ocw.mit.edu/courses/18-650-statistics-for-applications-fall-2016/
 pages/assignments/](https://ocw.mit.edu/courses/18-650-statistics-for-applications-fall-2016/pages/assignments/)

```
(%i6) load(draw)$
set_draw_defaults(line_width=2, grid = [2,2], point_type = filled_circle,
  head_type = 'nofilled, head_angle = 20, head_length = 0.5,
  background_color = light_gray, draw_realpart=false)$
load (descriptive)$ load (distrib)$
fpprintprec : 5$ ratprint : false$
```

Homemade functions fill, head, tail, Lsum are useful for looking at long lists.

```
(%i10) fill ( aL) := [ first (aL), last (aL), length (aL) ]$
head(L) := if listp (L) then rest (L, - (length (L) - 3) ) else
  error("Input to 'head' must be a list of expressions ")$
tail (L) := if listp (L) then rest (L, length (L) - 3 ) else
  error("Input to 'tail' must be a list of expressions ")$
Lsum (aList) := apply ("+", aList)$
```

3 *fracData (List, a, b)*

fracData (alist, a, b) calculates the fraction of the list numbers which lie in the interval [a, b].

```
(%i11) fracData (myData, xx1, xx2) :=
  block ([ ccnt : 0 ],
    for j thru length (myData) do
      if myData[j] >= xx1 and myData[j] <= xx2 then ccnt : ccnt + 1,
    float (ccnt/ length (myData)) )$
```

4 *confidence (q, m, s)*

With q a number in the interval $0 < q < 1$, and with m the mean and s the standard deviation of a Normal distribution, confidence (q, m, s) prints out the values dx, m - dx, m + dx, and outputs a list [m - dx, m + dx] which allows one to have 100*q % confidence a random value of x will lie within that interval, ie., within m +/- dx.

```
(%i12) confidence (qq, mm, ss) :=
  block ([ddx],
    ddx :
      float ( (quantile_normal (qq + (1 - qq)/2, mm, ss) -
        quantile_normal ( (1 - qq)/2 , mm, ss))/2 ),
    print ( "delx = ", ddx ),
    print ( " x1 = ", mm - ddx, ", x2 = ", mm + ddx ),
    [mm - ddx, mm + ddx])$
```

5 *Basic Concepts in Error Analysis*

[HH] Sec. 1.1

"The aim of error analysis is to quantify and record the errors associated with the inevitable spread in a set of measurements, and to identify how we may improve the experiment. In the physical sciences experiments are often performed in order to determine the value of a quantity. However, there will always be an error associated with that value due to experimental uncertainties."

"The sources of these uncertainties are discussed later in this chapter. We can never be certain what the exact value is, but the errors give us a characteristic range in which we believe the correct value lies with a specified likelihood."

[HH] Sec. 1.2.1:

"A precise measurement is one in which the spread σ of the results is 'small', either in absolute magnitude or relative to the accepted value."

"An accurate measurement is one in which the results of the experiment agree with the accepted value."

"Types of errors:

- 1.) random errors—these influence precision;
- 2.) systematic errors—these influence the accuracy of a result;
- 3.) mistakes—bad data points."

"The signature of random errors in an experiment is that repeated measurements are scattered over a range. The smaller the random uncertainty, the smaller the scattered range of the data. ...the spread of the data is defined by the standard deviation σ ."

"There are two categories of scatter in experiments: (1) technical, and (2) fundamental noise."

5.1 Spring-loaded Projectile Launcher Experiment

We will illustrate the results of an experiment measuring the range of a ball-bearing launched from a spring-loaded projectile launcher. Successive measurements were taken of the distance the ball-bearing projectile landed with respect to the launching cannon. We show a histogram of the projectile radial range, $r = \sqrt{x^2 + y^2}$, below. The crucial point to note is this: although nominally these repeated measurements are performed under exactly the same conditions, and Newton's laws of motion, which govern the trajectory, are time independent, successive repetitions of the experiment gave different values for the radial range of the ball-bearing projectile.

The histogram shown below contains the information about the range of the projectile, and the uncertainty in this range.

The Maxima function `random_normal (m,s)` returns a random variate, with $s > 0$. m is the mean and s the standard deviation. Calling `random_normal` with a third argument n , `random_normal (m,s,n)`, a random sample of size n will be simulated.

```
(%i13) for j thru 5 do print ( [random_normal (15,2), random_normal (7.5, 1.5) ] )$
```

```
[ 13.324,6.2441 ]
```

```
[ 17.594,6.3933 ]
```

```
[ 12.974,6.9634 ]
```

```
[ 13.112,8.7017 ]
```

```
[ 14.244,6.7052 ]
```

```
(%i14) nsample : 50$
```

```
(%i18) xvals : random_normal (15, 5, nsample)$
```

```
fill (xvals);
```

```
head (xvals);
```

```
tail (xvals);
```

```
(%o16) [ 22.594,9.5975,50 ]
```

```
(%o17) [ 22.594,16.226,12.055 ]
```

```
(%o18) [ 11.587,12.847,9.5975 ]
```

```
(%i22) yvals : random_normal (7.5, 3, nsample)$
```

```
fill (yvals);
```

```
head (yvals);
```

```
tail (yvals);
```

```
(%o20) [ 10.449,9.883,50 ]
```

```
(%o21) [ 10.449,5.324,6.3835 ]
```

```
(%o22) [ 8.4039,11.276,9.883 ]
```

```
(%i26) xyhits : makelist ( [ xvals [j], yvals [j] ], j, 1, nsample)$
```

```
fill (xyhits);
```

```
head (xyhits);
```

```
tail (xyhits);
```

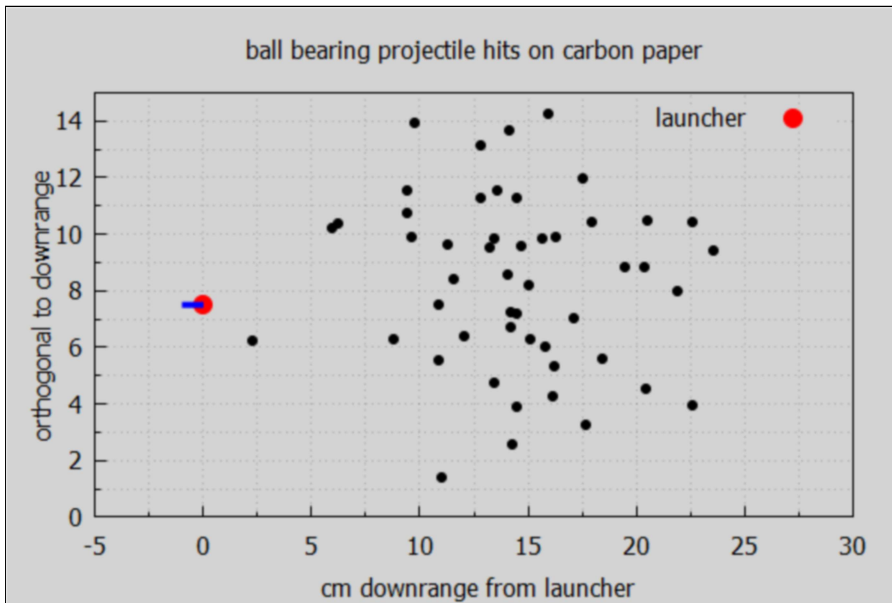
```
(%o24) [ [ 22.594,10.449 ], [ 9.5975,9.883 ], 50 ]
```

```
(%o25) [ [ 22.594,10.449 ], [ 16.226,5.324 ], [ 12.055,6.3835 ] ]
```

```
(%o26) [ [ 11.587,8.4039 ], [ 12.847,11.276 ], [ 9.5975,9.883 ] ]
```

```
(%i27) wxdraw2d ( xrange = [-5,30], yrange = [0, 15], color = black,
  title = "ball bearing projectile hits on carbon paper",
  xlabel = " cm downrange from launcher",
  ylabel = "orthogonal to downrange",
  points (xyhits), color = red, key = "launcher",
  point_size = 2, points ( [[0, 7.5] ] ), color = blue, key = "",
  line_width = 4, explicit (7.5, x, -1, 0) )$
```

(%t27)



```
(%i31) radialRange : makelist ( sqrt ( xvals[j]^2 + yvals[j]^2 ), j, 1, nsample)$
fill (radialRange);
head (radialRange);
tail (radialRange);
```

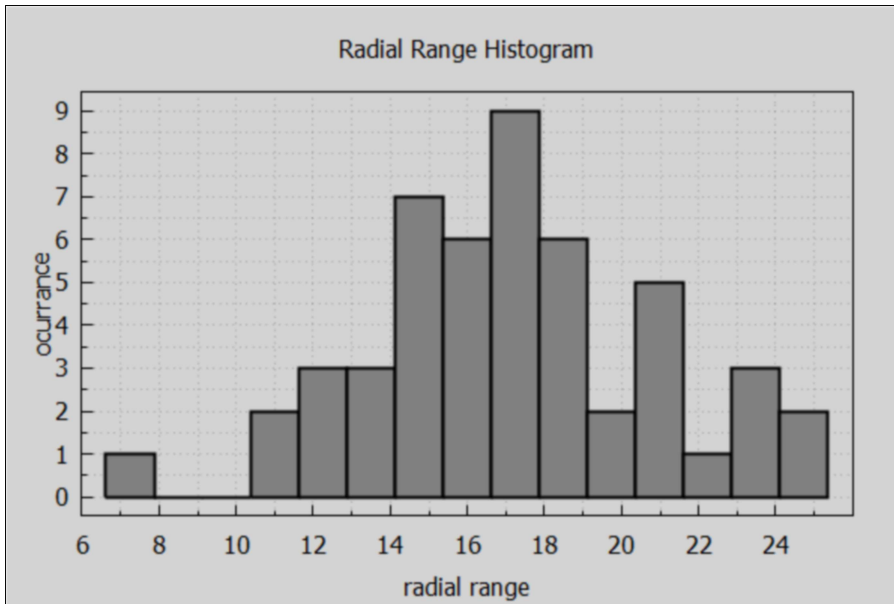
(%o29) [24.893, 13.776, 50]

(%o30) [24.893, 17.077, 13.641]

(%o31) [14.314, 17.094, 13.776]

```
(%i32) wxhistogram (radialRange, xlabel = "radial range", ylabel = "ocurrance",
title = " Radial Range Histogram ", fill_color = black,
fill_density = 0.5 /* , frequency = density */ , nclasses = 15 )$
```

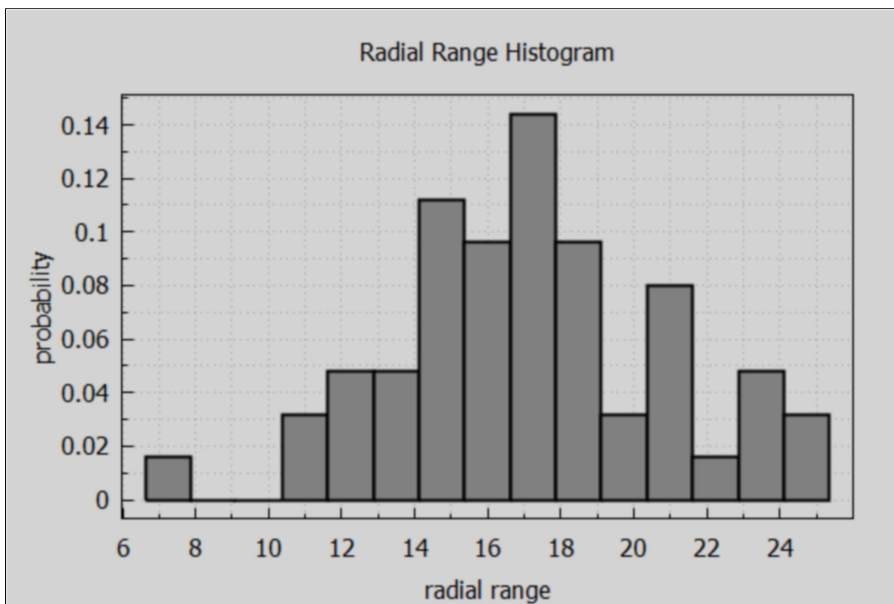
(%t32)



If we had included the option frequency = density, we would get automatic normalization so the area under the bars would equal unity and the y axis would extend from 0 to 1, and we would label the y axis 'probability'. The default number of bins is 10. Using the option nclasses = 15, as we did above, throws the data into 15 bins.

```
(%i33) wxhistogram (radialRange, xlabel = "radial range", ylabel = "probability",
title = " Radial Range Histogram ", fill_color = black,
fill_density = 0.5, frequency = density, nclasses = 15 )$
```

(%t33)



```
(%i35) mm : mean (radialRange);  
        ss : std (radialRange);  
(mm)   17.11  
(ss)   3.7959
```

The mean (average) value of the radialRange values is mm cm and one standard deviation is ss cm. We have used the Maxima functions mean (data) and std (data). We have previously used both of these functions in Stat04-Normal.wmxm, in the section 'Random Sample Size n = 100 Simulations'. The best estimate of the measured quantity is the mean $\langle x \rangle$ of the distributed data.

```
(%i36) fracData (radialRange, mm - ss, mm + ss);  
(%o36) 0.66
```

The value returned by fracData above is the fraction of the radialRange values which lie in the range [mean - std, mean + std], ie., within one standard deviation of the mean value.

6 *Precision and Accuracy Defined*

We either summarize or quote material in [HH], Sec. 1.2.

There is a great difference between saying "this is an accurate measurement of x" or saying "this is a precise measurement of x".

A 'precise measurement' means the spread of the experimental values is 'small', either small in absolute value or the ratio: $\sigma/\langle x \rangle$ is small. The spread of the data is defined by the standard deviation σ .

An 'accurate measurement' means that the mean value $\langle x \rangle$ is in 'agreement' with the current accepted value. Obviously, we could not use the word 'accurate' to describe the measurement of some quantity whose value is unknown so far, either experimentally or theoretically.

1.) An experiment which is both precise and accurate means the spread of the measurements is small and the mean of the measurements agrees with the accepted value. This is the 'gold standard'.

2.) An experiment which is accurate but not precise means that the mean of the measurements agrees with the accepted value, but the spread of the measured values is large.

3.) An experiment which is precise but not accurate has a small spread of measurements (small random errors) and has some systematic error which causes the mean of the measurements to be 'far' from the accepted value.

4.) An experiment which in neither precise nor accurate characterizes a set of measurements which have both a large systematic error or errors and well as large random errors: the spread of measured values is large and the mean is 'far' from the accepted value.

6.1 Systematic Errors

Quoting [HH] in Sec. 1.2.3:

"Systematic errors cause the measured quantity to be shifted away from the accepted, or predicted, value. Measurements where this shift is small (relative to the error) are described as accurate. For example, for the data from the projectile launcher shown" above, " the total range needs to be measured from the initial position of the ball-bearing within the launcher. Measuring from the end of the cannon produces a systematic error. Systematic errors are reduced by estimating their possible size by considering the apparatus being used and observational procedures."

6.2 Mistakes

"Another class of error which defies mathematical analysis is a mistake. These are similar in nature to systematic errors, and can be difficult to detect. Writing 2.34 instead of 2.43 in a lab book is a mistake, and if not immediately corrected is very difficult to compensate for later."

6.2.1 20% of measurements incorrectly logged example

"A histogram of events is shown below, where the automated data-collecting software misfired on 20% of the events. Malfunction of the apparatus can be difficult to spot; the presence of erroneous points can become apparent when the data are displayed graphically."

(%i39) n1 : 800\$
 n2 : 200\$
 nn : n1 + n2\$

n1 events are correctly logged.

```
(%i43) xv1 : random_normal (15, 2, n1)$
      yv1 : random_normal (7.5, 1, n1)$
      xy1 : makelist ([ xv1 [j], yv1 [j] ], j, 1, n1)$
      fill (xy1);
```

```
(%o43) [[13.548,7.4],[11.999,7.4676],800]
```

n2 events are not correctly logged.

```
(%i47) xv2 : random_normal (8, 2, n2)$
      yv2 : random_normal (7.5, 1, n2)$
      xy2 : makelist ([ xv2 [j], yv2 [j] ], j, 1, n2)$
      fill (xy2);
```

```
(%o47) [[6.9772,7.7799],[6.7422,6.8389],200]
```

We can use the Maxima function join (list1, list2) to 'merge' two lists of the same length.

We can use the Maxima function append (list1, list2) to 'merge' two lists of any length.

```
(%i51) L1 : [a, b, c];
      L2 : [z, y, x, w, v];
      LLjoin : join ( L1, L2 );
      LLappend : append ( L1, L2 );
```

```
(L1) [a,b,c]
```

```
(L2) [z,y,x,w,v]
```

```
(LLjoin) [a,z,b,y,c,x]
```

```
(LLappend) [a,b,c,z,y,x,w,v]
```

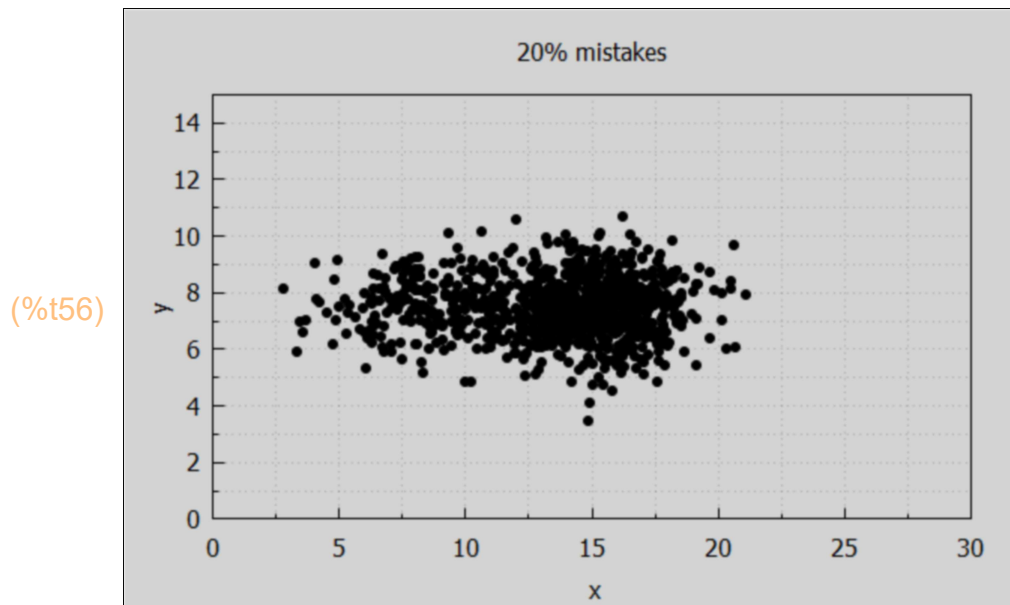
```
(%i55) xytotal : append (xy1, xy2)$
      fill (xytotal);
      head (xytotal);
      tail (xytotal);
```

```
(%o53) [[13.548,7.4],[6.7422,6.8389],1000]
```

```
(%o54) [[13.548,7.4],[16.392,7.8997],[14.807,9.4778]]
```

```
(%o55) [[7.6779,8.1847],[8.3842,7.4179],[6.7422,6.8389]]
```

```
(%i56) wxdraw2d ( xrange = [0,30], yrange = [0, 15], color = black,
                title = "20% mistakes",
                xlabel = "x", ylabel = "y", points (xytotal) )$
```



To construct a histogram, we need to create a list from the raw data which is a simple list of scalars. This could have the form:

```
rx : makelist ( sqrt ( xvtot[j]^2 + yvtot[j]^2 ), j, 1, nsample)$
```

```
(%i60) xvtot : append (xv1, xv2)$
      fill (xvtot);
      yvtot : append (yv1, yv2)$
      fill (yvtot);
```

```
(%o58) [ 13.548, 6.7422, 1000 ]
```

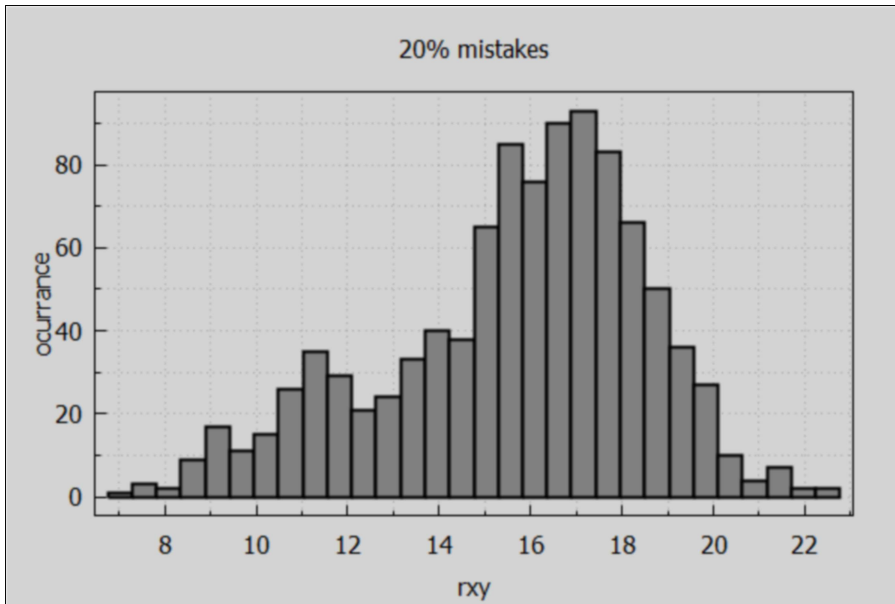
```
(%o60) [ 7.4, 6.8389, 1000 ]
```

```
(%i62) rxy : makelist ( sqrt ( xvtot[j]^2 + yvtot[j]^2 ), j, 1, nn)$
      fill (rxy);
```

```
(%o62) [ 15.437, 9.6036, 1000 ]
```

```
(%i63) wxhistogram (rxy, xlabel = "rxy", ylabel = "ocurrence",
title = " 20% mistakes ", fill_color = black,
fill_density = 0.5 /* , frequency = density */ , nclasses = 30 )$
```

(%t63)



The bump down near rxy = 11 indicates the possibility of incorrect logging of raw data by automatic data logging equipment.

6.2.2 Other Examples of Mistakes

"There are other well-known types of mistakes which can influence the precision and accuracy of experimental results with potentially disastrous consequences. Misreading scales occurs often with an analog device which has a 0–10 scale above the gradations, and a 0–3 scale underneath. Care has to be taken when using a signal generator where an analog dial from 1 to 10 is used in conjunction with multiplier knobs such as 1–10 kHz. With an instrument such as an oscilloscope one has to be careful to check whether a multiplier such as $\times 10$ has been engaged."

"There are many examples where confusion over units has had disastrous consequences. In 1999 the failure of the NASA Mars Climate Orbiter was attributed to confusion about the value of forces: some computer codes used SI units, whereas others used imperial."

"A Boeing 767 aircraft ran out of fuel mid-flight in 1983; a subsequent investigation indicated a misunderstanding between metric and imperial units of volume."

6.3 Rules of Thumb for Precision

We either summarize or quote [HH], Sec. 1.3.

- "1.) The precision of a measurement only equals the precision of the measuring device when all repeated measurements are identical.
- 2.) The highest precision achievable with an analog device such as a ruler is half a division.
- 3.) The precision of a digital meter is limited to the last decimal point; i.e. one in the last digit."

6.4 Achieving Accurate Measurements

We either summarize or quote [HH], Sec. 1.4.

"The accuracy of an experiment is determined by systematic errors. For the example of the projectile launcher, the range is a function of both the launch angle and muzzle speed. Experimental factors which could affect the accuracy include the setting of the launch angle or the reliance on the manufacturer's specification of the muzzle speed. The theoretical prediction, for this example, is based on a calculation which ignores air resistance—the validity of this assumption could be questioned."

"Three of the more common sources of systematic error are zero, calibration, and insertion errors.

- 1.) An example of a zero error is using a ruler to measure length if the end of the ruler has been worn away.
- 2.) An example of a calibration error is using a metal ruler calibrated at 20 deg C for a measurement when the temperature is 10 deg C, ignoring the thermal contraction.
- 3.) Examples of insertion errors include:
 - a.) placing a room-temperature thermometer in a hot fluid, which will change the temperature of the hot fluid,
 - b.) the current in an electrical circuit being changed by placing a meter across a component of the circuit."

"In general one should ensure the apparatus used is properly calibrated and zeroed."

7 *Measurement Errors and the Central Limit Theorem*

Quoting from R. J. Barlow, Sec. 4.1

"When values are quoted with an error, this error is a Gaussian standard deviation σ . If you say the length of a piece of string is 12.3 ± 0.1 cm, you mean that you have measured it with a ruler (or equivalent device) which gives answers that differ from the true value by within 0.1 cm 68% of the time, 0.2 cm 95% of the time, and 0.3 cm 99.7% of the time..."

"This is not just an arbitrary choice. Errors on measurements and results are generally well described by the Gaussian [Normal] distribution, which is of course why it is also known as the normal distribution."

7.1 Why Errors are Gaussian

Quoting Bohm, Sec. 4.1

"When we talk about measurement errors, we do not mean mistakes caused by the experimenter, but the unavoidable random dispersion of measurements. Therefore, a better name would be measurement uncertainties. We will use the terms uncertainty and error synonymously."

"The natural sciences owe their success to the possibility to compare quantitative hypotheses to experimental facts. However, we are able to check theoretical predictions only if we have an idea about the accuracy of the measurements. If this is not the case, our measurements are completely useless."

"Of course, we also want to compare the results of different experiments to each other and to combine them. Measurement errors must be defined in such a way that this is possible without knowing details of the measurement procedure. Only then, important parameters, like constants of nature, can be determined more and more accurately and possible variations with time, like it was hypothesized for the gravitational constant, can be tested."

"Finally, it is indispensable for the utilization of measured data in other scientific or engineering applications to know their accuracy and reliability. An overestimated error can lead to a waste of resources and, even worse, an underestimated error may lead to wrong conclusions."

7.2 Expectation $E(F)$ and Variance $V(F)$ Defined

7.2.1 $E(F)$

The expectation $E(F)$, or $\langle F \rangle$, of a function $F(x_1, x_2, \dots, x_N)$ is an N dimensional integral

$$\int \int \dots \int F(x_1, x_2, \dots, x_N) * P(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N$$

with integration over the whole domain of the variables x_1, x_2, \dots, x_N , usually minus infinity (minf) to positive infinity (inf).

$P(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N$ is the joint probability that x_1 will be found in the interval $[x_1, x_1 + dx_1]$, x_2 will be found in the interval $[x_2, x_2 + dx_2]$, ..., x_N will be found in the interval $[x_N, x_N + dx_N]$. The joint probability P has the normalization

$$\int \int \dots \int P(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N = 1.$$

The random variables x_1, x_2, \dots, x_N are independent if and only if P factors:

$$P(x_1, x_2, \dots, x_N) = P_1(x_1) * P_2(x_2) * \dots * P_N(x_N).$$

Then we require the normalization conditions:

$$\int P_1(x_1) dx_1 = \int P_2(x_2) dx_2 = \dots = \int P_N(x_N) dx_N = 1.$$

7.2.2 $V(F)$

The variance of some function $F(x_1, x_2, \dots, x_N)$ is denoted $V(F)$ and defined by

$$V(F) = E([F - E(F)]^2)$$

which, in words, is the expectation of the square of the difference between F and the expectation of F .

7.2.3 $\sigma(F)$

The standard deviation σ is the square root of the variance.

7.3 $E(a_1 x_1 + a_2 x_2) = a_1 E(x_1) + a_2 E(x_2)$

Let's show that any linear combination of the x_i , such as

$$X = \sum a_i x_i, \text{ with } i \text{ in } [1, N], \text{ satisfies:}$$

$$E(X) = \langle X \rangle = \text{mean of } X = \sum a_i E(x_i),$$

that is, the mean of X is the same linear combination of the means of the x_i .

Working through the case $N = 2$ will make clear the generality of the result, which doesn't depend on the factorization of $P(x_1, x_2)$ (ie., doesn't depend on x_1 and x_2 being independent random variables).

The mean of a general function of x_1 and x_2 , $F(x_1, x_2)$ is defined as (if x_1 and x_2 are real variables; we will assume they can take values between minus infinity ("minf") and positive infinity ("inf")).

$$E(F(x_1, x_2)) = \int \int F(x_1, x_2) P(x_1, x_2) dx_1 dx_2,$$

where $P(x_1, x_2)$ is the joint probability that x_1 and x_2 have particular values, with the normalization requirement:

$$\int \int P(x_1, x_2) dx_1 dx_2 = 1.$$

The expectation values for x_1 and x_2 are (in general):

$$E(x_1) = \int \int x_1 P(x_1, x_2) dx_1 dx_2,$$

$$E(x_2) = \int \int x_2 P(x_1, x_2) dx_1 dx_2.$$

Now assume $F(x_1, x_2)$ is a linear function of independent variables x_1 and x_2 :

$$F(x_1, x_2) = a_1 x_1 + a_2 x_2, \text{ with } a_1 \text{ and } a_2 \text{ independent of } x_1 \text{ and } x_2. \text{ Then:}$$

$$\begin{aligned} E(a_1 x_1 + a_2 x_2) &= \int \int (a_1 x_1 + a_2 x_2) P(x_1, x_2) dx_1 dx_2, \\ &= a_1 \int \int x_1 P(x_1, x_2) dx_1 dx_2 + a_2 \int \int x_2 P(x_1, x_2) dx_1 dx_2 \\ &= a_1 E(x_1) + a_2 E(x_2). \text{ QED.} \end{aligned}$$

In the general case,

$$E(\sum a_i x_i) = \sum a_i E(x_i).$$

7.4 $V(a_1 x_1 + a_2 x_2) = (a_1)^2 V(x_1) + (a_2)^2 V(x_2)$

In a similar manner one can prove (James, p. 28, Eq. (2.32)) the relation between the variances is (with x_1 and x_2 independent variables, a_1 and a_2 constants):

$$\begin{aligned} V(a_1 x_1 + a_2 x_2) &= (a_1)^2 V(x_1) + (a_2)^2 V(x_2), \text{ and generalizing to arbitrary } N, \\ V(\sum a_i x_i) &= \sum (a_i)^2 V(x_i). \end{aligned}$$

The above relations hold even if the independent variables x_i are each described by completely different probability distributions $P_i(x_i)$.

We prove this for the case $N = 2$.

With $F = a_1 x_1 + a_2 x_2$, $V(F) = E([F - E(F)]^2)$ by definition.

$$E(F) = a_1 E(x_1) + a_2 E(x_2) = a_1 \mu_1 + a_2 \mu_2.$$

$$F - E(F) = a_1 (x_1 - \mu_1) + a_2 (x_2 - \mu_2).$$

$$[F - E(F)]^2 = (a_1)^2 (x_1 - \mu_1)^2 + (a_2)^2 (x_2 - \mu_2)^2 + 2 a_1 a_2 (x_1 - \mu_1) (x_2 - \mu_2).$$

$$E([F - E(F)]^2) = (a_1)^2 E((x_1 - \mu_1)^2) + (a_2)^2 E((x_2 - \mu_2)^2) + 2 a_1 a_2 E((x_1 - \mu_1)(x_2 - \mu_2)).$$

The first two terms on the rhs are:

$$(a_1)^2 V(x_1) + (a_2)^2 V(x_2).$$

In the third term, because x_1 and x_2 are independent random variables, $P(x_1, x_2) = P_1(x_1) P_2(x_2)$, and the third term is proportional to:

$$\begin{aligned} E[(x_1 - \mu_1)(x_2 - \mu_2)] &= \int \int (x_1 - \mu_1)(x_2 - \mu_2) P(x_1, x_2) dx_1 dx_2, \\ &= \int \int (x_1 - \mu_1)(x_2 - \mu_2) P_1(x_1) P_2(x_2) dx_1 dx_2 \\ &= \int (x_1 - \mu_1) P_1(x_1) dx_1 \int (x_2 - \mu_2) P_2(x_2) dx_2 \\ &= \left[\int x_1 P_1(x_1) dx_1 - \mu_1 \right] \left[\int x_2 P_2(x_2) dx_2 - \mu_2 \right] \\ &= (\mu_1 - \mu_1)(\mu_2 - \mu_2) = 0. \end{aligned}$$

Thus $V(a_1 x_1 + a_2 x_2) = (a_1)^2 V(x_1) + (a_2)^2 V(x_2)$, which can be generalized to arbitrary values of N .

7.5 Central Limit Theorem (CLT)

Quoting from R. J. Barlow, Sec. 4.1

"Measurements acquire errors from many different sources. If you measure the length of a rod using a ruler, all sorts of inexactitudes creep in: optical parallax, the ruler's calibration, rounding errors, your hand shaking, and so on. Reading a meter with a moving pointer has similar problems. Digital meters and electronic readout avoid the effects of shaky hands and bleary eyes, but at the expense of others, hidden in the depths of the electronics. The imperfections in the measurements we make are not due to one cause, but to many."

"Now, there is a powerful and surprising result about the behaviour of a variable which is the sum of several others. It is called the central limit theorem or CLT for short."

"The central limit theorem:

If you take $X = \sum x_i$, where $i = 1, 2, 3, \dots, N$, with each x_i a measurement of some property x taken from a distribution of mean μ_i and variance $V_i = (\sigma_i)^2$, then the distribution for X :

(a) has an expectation value

$$E(X) = \langle X \rangle = \sum \mu_i / N \quad (4.1)$$

(b) has a variance

$$V(X) = \sum V_i / N == \sum (\sigma_i)^2 / N \quad (4.2)$$

(c) becomes Gaussian as N approaches infinity. (4.3) "

"This is why the Gaussian is so important. A quantity produced by the cumulative effect of many independent variables will be, at least approximately, Gaussian, no matter what the distributions of the original variables may have been. Measurement errors behave accordingly, as do many other observed quantities. For example, human heights are well described by a Gaussian distribution, as are the lengths of the arm, forefinger, and other anatomical measurements, as these are due to the combined effects of many genetic and environmental factors."

7.6 Justification of the Central Limit Theorem

Suppose you take N measurements, each done under the same conditions, and conclude your "best" value is the arithmetic average of the N values, call it x_{av} . (In texts this usually is denoted by x with an overline, or " x -bar"). If you then repeat the whole process with another N measurements, you will, in general, get a somewhat different value for x_{av} , so x_{av} itself will have some distribution of values, with a mean $\langle x_{av} \rangle$ and its own variance $V(x_{av}) = (s_{av})^2$.

Now with x_i standing for the i 'th measurement of property x , out of a total of N measurements in a given experiment, let:

$$x_{av} = \sum x_i / N = \text{average of } N \text{ measurements.}$$

If the same experiment, with N measurements of property x carried out with the same experimental conditions, the value of x_{av} will, in general, be slightly different. What can we say about the average of the set of x_{av} 's? Let $\langle x_{av} \rangle$ designate the average of the set of x_{av} 's.

Using our relations for a linear combination of random variables x_i ,

$$\langle x_{av} \rangle = E(x_{av}) = \sum E(x_i) / N$$

If we assume that $E(x_i) \sim \mu$ for all i , we get

$$\langle x_{av} \rangle = E(x_{av}) \sim \sum \mu / N = (N \mu) / N = \mu.$$

" x_{av} is subject to random statistical fluctuations, but on average its value is μ . The difference between your actual measured x_{av} 's and the 'true' value μ is described by some distribution which has variance $V(x_{av}) = (s^2 / N)$ [in which s is approximately one standard deviation (the uncertainty) of each measurement], as we show below."

Next consider the variance of x_{av} and use our relation for the variance of a linear combination of variables and also our assumption that each of the N measurements of x in each experiment are independent (the result of one careful measurement has no influence on a second careful measurement).

$V(x_{av}) = V(\sum x_i / N) = V(\sum (x_i / N)) = (\sum V(x_i)) / N^2$, in which we have used the variance of a linear combination of random variables relation proved above, namely that $V(a_1 x_1 + a_2 x_2 + \dots + a_n x_n) = a_1^2 V(x_1) + a_2^2 V(x_2) + \dots + a_n^2 V(x_n)$, in a situation in which $a_1 = (1/N)$, $a_2 = (1/N)$, etc., so every term here has a $(1/N)^2$ coefficient which can be factored out of the sum.

If we assume that the $V(x_i)$ all have the same value s^2 [since each measurement has the same uncertainty and hence the same s , we get

$$V(x_{av}) = (\sum s^2) / N^2 = (N s^2) / N^2 = s^2 / N = s_{av}^2, \text{ as asserted above.}$$

Taking the square root:

$$s_{av} = s / \sqrt{N} = s / N^{(1/2)}.$$

The spread of the distribution of values of the x_{av} 's will be smaller than s , depending on the size of N . (Averaging is good for you. More measurements are better than fewer.)

" s_{av}/\sqrt{N} is called the 'standard error of the mean', as it describes how well you know the mean of a distribution, which is often a very important quantity."

Common symbols are: SE, SEM, σ_M , σ_E . a statistic that indicates how much the average value (mean) for a particular sample is likely to differ from the average value for the larger population from which it is drawn.

7.7 Photon Decays

Continuing with Barlow's examples:

"Example: Photon energy measurements The energy resolution of a γ -ray detector used to investigate a decaying nuclear isotope is 50 keV. If only one such decay is observed, its energy is known to 50 keV. If 100 are collected, this improves to 5 keV. To reach 1 keV you would need to observe 2500 decays."

In this example, the uncertainty of an individual measurement is $s = 50$ keV (50 kiloVolts). If this measurement produces a value $E = (300 \pm 50)$ keV based on one decay, we expect that averaging 100 measurements (with the identical equipment and setup) will produce an answer with an uncertainty of about $50/10 = 5$ keV. Averaging 2500 measurements will produce an answer with an uncertainty of about $50 / 50 = 1$ keV.

7.8 Weights of Eggs

"Example Weights of eggs The weights of eggs produced by a farmer's hens have a standard deviation of 10 g. He feeds a group of hens an expensive vitamin supplement, which will pay back its costs if it increases the weight of the eggs by 2 g. He measures 25 eggs from vitamin-fed hens and their average has increased by 3 g. Does this prove anything useful?"

"No. The standard error on the mean is $s/\sqrt{N} = 10/\sqrt{25} = 2$ g, so the increase [in weight] is only $1.5 s_{\text{av}}$ and is not really significant."

"Notice the subtle difference in this example, in that the variation in the egg weights comes from the spread in the objects themselves, not from the measurement process; it is assumed that the eggs can be weighed with complete accuracy, or at any rate an accuracy much better than 10 g, which is probably true."

Chicken eggs vary in weight from about 60 g to 70 g, depending on the age of the chicken.