Propagation of Error

Keegan Korthauer

Department of Statistics

UW Madison

Error Propagation

- Error is inherent in any measuring procedure
- Measured values will differ from the underlying (often unknown) true value
- Calculations performed on these measurements will also contain error (i.e. they are **propagated** from the measurement to the calculated value)

Error in Measurements
e.g. length and width of a rectangle

→ **Calculation using Measurements**
e.g. length x width

→ **Error in Calculated Value**
e.g. area of a rectangle

- There are ways to estimate the likely size of the error in a calculation from the size of the measurement error

MEASUREMENT ERROR

Bias vs Random Error

Estimating Uncertainty

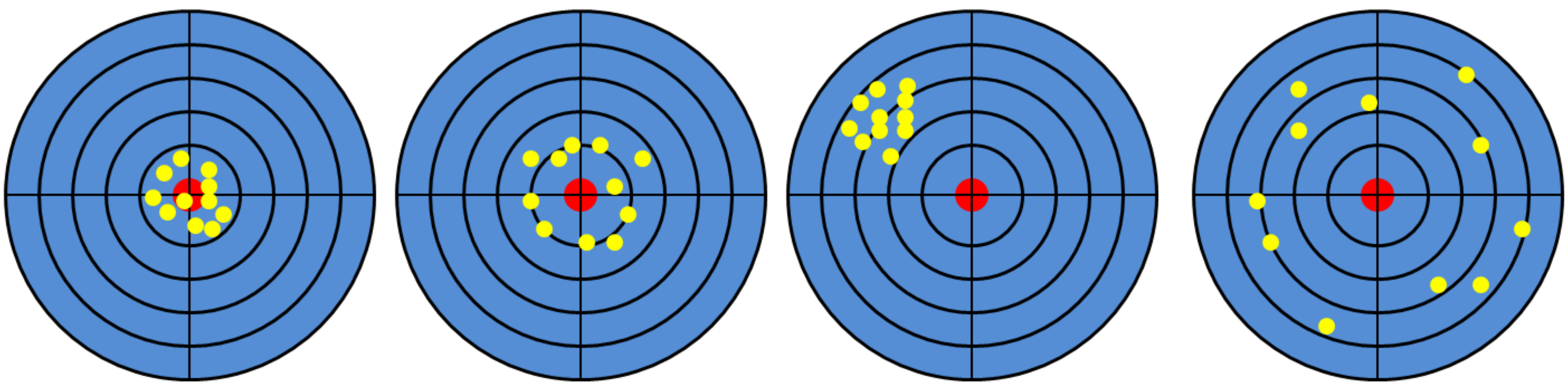
Bias and Random Error

$$\text{Measured Value} = \text{True Value} + \underbrace{\boxed{\text{Bias} + \text{Random Error}}}_{\text{Total Error}}$$

- Bias (also known as Systematic Error)
  - Same for every measurement
  - Example: lab tech forgets to tare scale when weighing product in a dish
- Random Error
  - Varies from measurement to measurement
  - Likely to average out to zero in the long run
  - Example: Rounding or interpolating between graduation marks

Accuracy and Precision

- **Accuracy** is how close the mean value μ is to the true value being measured (determined by bias)
- **Precision** is the degree to which repeated measurements tend to agree with each other (determined by variance)



Accurate & Precise

Accurate & Imprecise

Inaccurate & Precise

Inaccurate & Imprecise

www.knowitsdifference.com

Summary

For a measured random variable with mean μ and standard deviation σ:

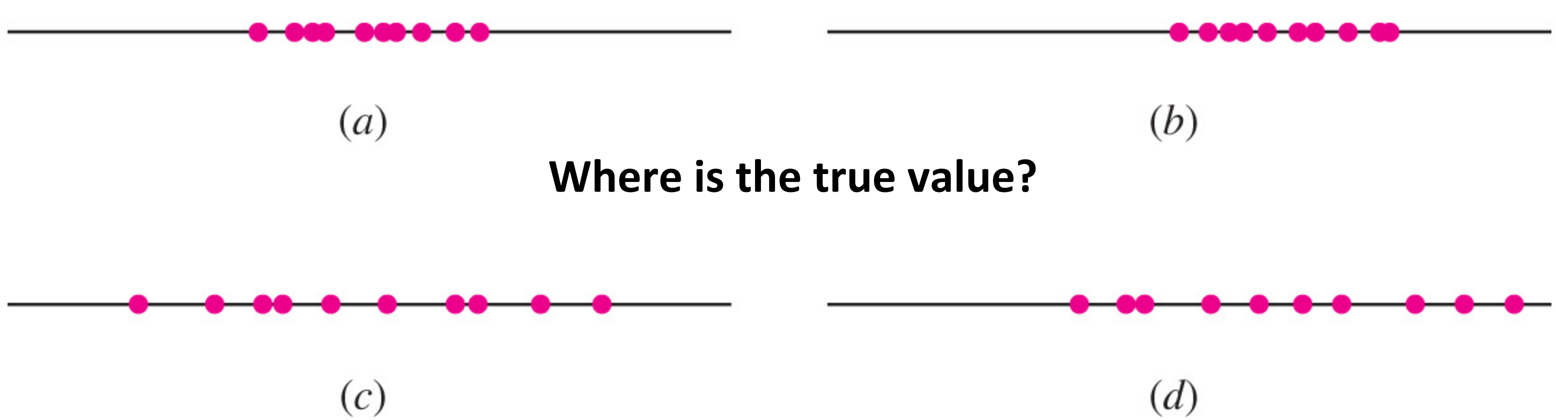
- **Bias** = μ – true value
- **Uncertainty** (size of typical random error) = σ
- **Accuracy** corresponds to low bias
- **Precision** corresponds to low uncertainty

Estimation of Bias

- An estimate of bias is the difference between the true value and the average of *repeated measurements of the same quantity*

$$\text{Bias} \approx \text{Sample Mean} - \text{True Value} = \bar{X} - \text{True Value}$$

- When we do not know the true value that we are measuring (usually the case), we cannot estimate bias



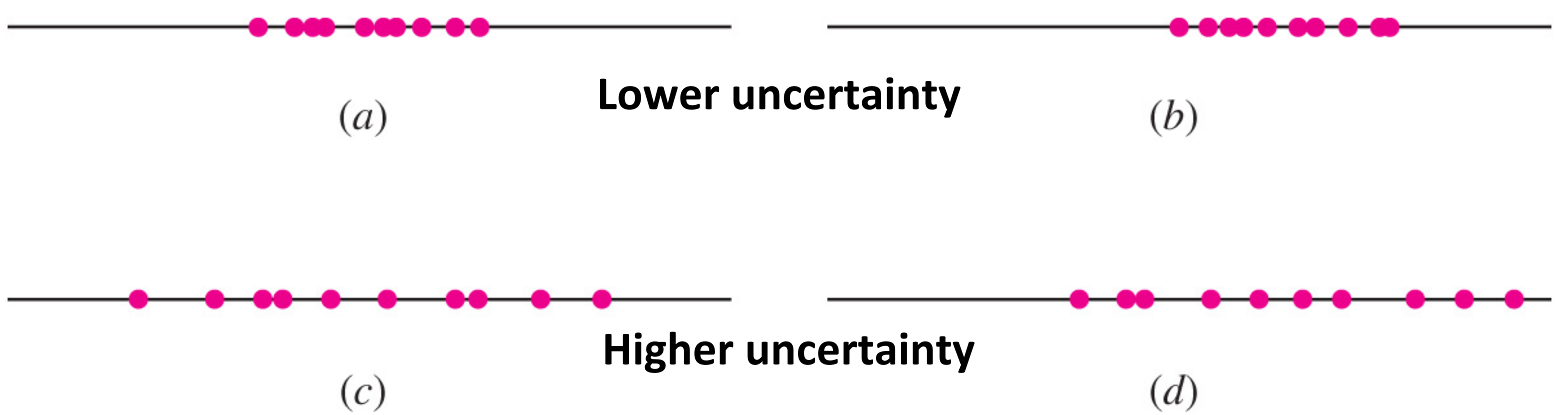
Where is the true value?

Estimation of Uncertainty

- We can estimate uncertainty as the standard deviation of *repeated measurements of the same quantity*

Uncertainty ≈ Sample Standard Deviation = σ

- It is best to have a large sample size for a good estimate of σ



Example – Exercise 3.1.7

- A replica of The Kilogram (the standard weight that represents exactly 1kg) is weighed five times
- The measurements are (in units of micrograms above 1kg):

114.3, 82.6, 136.4, 126.8, 100.7

(a) Can we estimate the uncertainty?

(b) Can we estimate the bias?

Example – Exercise 3.1.7

- A replica of The Kilogram (the standard weight that represents exactly 1kg) is weighed five times
- The measurements are (in units of micrograms above 1kg):

114.3, 82.6, 136.4, 126.8, 100.7

(a) Can we estimate the uncertainty?

Yes, estimate with sample sd = 21.3 micrograms

(a) Can we estimate the bias?

No, since we do not know the true value

UNCERTAINTY OF A LINEAR COMBINATION OF MEASUREMENTS

Linear Combinations of Independent Repeated Measurements

Linear Combinations of Dependent Repeated Measurements

This should look familiar…

- How do we estimate uncertainty for a measurement multiplied by a constant? (e.g. unit conversion)
- How do we estimate uncertainty for the sum of two or more measurements? (e.g. adding commute times)
- Measurements are **random variables** so we can apply what we learned in Chapter 2, section 5 about finding variances of linear combinations of random variables

Summary of Uncertainty for Linear Combinations of Measurements

- Let $X_1,\dots,X_n$ be **independent** measurements and $c_1,\dots,c_n$ be a constants. Then,

$$\mathrm{sd}(c_1X_1+\dots+c_nX_n)=\sigma_{c_1X_1+\dots+c_nX_n}=\sqrt{c_1^2\sigma_{X_1}^2+\dots+c_n^2\sigma_{X_n}^2}$$

- If $X_1,\dots,X_n$ are **dependent**, then we can only get an upper bound on the uncertainty of the linear combination

$$\mathrm{sd}(c_1X_1+\dots+c_nX_n)=\sigma_{c_1X_1+\dots+c_nX_n}\le|c_1|\sigma_{X_1}+\dots+|c_n|\sigma_{X_n}$$

Example 3.6

Suppose a surveyor needs to know the perimeter of a rectangular lot and takes repeated measures of two adjacent sides and finds that side X has a mean of 50.11m and an sd of 0.05m. Side Y has a mean of 75.21m with an sd of 0.08m. Assume measurements X and Y are independent.

What is the uncertainty in the estimate of the perimeter?

Example 3.6

Suppose a surveyor needs to know the perimeter of a rectangular lot and takes repeated measures of two adjacent sides and finds that side X has a mean of 50.11m and an sd of 0.05m. Side Y has a mean of 75.21m with an sd of 0.08m. Assume measurements X and Y are independent.

Perimeter Estimate: $P = 2X + 2Y = 250.64\text{m}$

The uncertainty in P is:

$$\sigma_P = \sigma_{2X+2Y} = \text{sqrt}(\sigma^2_{2X+2Y})$$

$$= \text{sqrt}(2^2\sigma^2_X + 2^2\sigma^2_Y)$$ <- (Since X and Y are independent)

$$= \text{sqrt}(4(0.05)^2 + 4(0.08)^2)$$

$$= 0.19\text{m}$$

Example 3.6 Continued

The surveyor's assistant came up with a different result. His solution:

Perimeter Estimate: P = 2X + 2Y = = X + X + Y + Y, so the uncertainty in P is:

$$\sigma_P = \sigma_{X+X+Y+Y} = \text{sqrt}(\sigma^2_{X+X+Y+Y})$$
$$= \text{sqrt}(\sigma^2_X + \sigma^2_X + \sigma^2_Y + \sigma^2_Y)$$
$$= \text{sqrt}((0.05)^2 + (0.05)^2 + (0.08)^2 + (0.08)^2)$$
$$= 0.13\text{m}$$

What happened here?

Example 3.6 Continued

The surveyor's assistant came up with a different result. His solution:

Perimeter Estimate: $P = 2X + 2Y = = X + X + Y + Y$, so the uncertainty in P is:

$$\sigma_P = \sigma_{X+X+Y+Y} = \text{sqrt}(\sigma^2_{X+X+Y+Y})$$

Assumes independence!

$$= \text{sqrt}(\sigma^2_X + \sigma^2_X + \sigma^2_Y + \sigma^2_Y)$$

$$= \text{sqrt}((0.05)^2 + (0.05)^2 + (0.08)^2 + (0.08)^2)$$

$$= 0.13\text{m}$$

What happened here?

Commute Time Example Revisited

- Let X_1 represent the time it takes (in minutes) to walk from my house to the bus stop. Assume $E(X_1)=3$, $Var(X_1)=1$.
- Let X_2 represent the time it takes the bus to travel between the bus stop and campus. Assume $E(X_2)=8$, $Var(X_2)=4$.
- What can we say about the uncertainty of $Y= X_1 + X_2$ if the two legs are NOT independent?

$$\sigma_{X_1+X_2} \leq |1|\sigma_{X_1} + |1|\sigma_{X_2}$$

$$= \sqrt{\sigma_{X_1}^2} + \sqrt{\sigma_{X_2}^2}$$

$$= \sqrt{4} + \sqrt{1} = 3$$

UNCERTAINTIES OF FUNCTIONS OF MEASUREMENTS

Functions of one measurement

Functions of several independent measurements

Functions of several dependent measurements

Uncertainty for a (Nonlinear) Function of a Measurement

- Recall that given a RV X and an arbitrary function U=U(X) we can find the mean of U(X) if we have its PMF (if discrete) or PDF (if continuous):

$$E[U(X)] = \sum_x U(x)p(x) \quad \text{or} \quad E[U(X)] = \int_{-\infty}^{\infty} U(x)f(x)\,dx$$

- How do we estimate the uncertainty of U(X), σ_U?
- If X is a measurement with uncertainty σ_X and σ_X is small, and if X is estimated by x, then

$$\sigma_U \approx \left| \left(\left. \frac{dU}{dX} \right|_{X=x} \right) \right| \sigma_X$$

Uncertainty for a (Nonlinear) Function of a Measurement

This equation is known as the **propagation of error** formula

$$\sigma_U \approx \left| \left(\frac{dU}{dX}\bigg|_{X=x} \right) \right| \sigma_X$$

Notes:

(a) it is only an approximation (see pgs 183-184 for derivation) [Taylor series approx of U ≈ constant + dU/dX*(X)]

(a) the estimate can be biased if either (1) σ_X is large or (2) the second derivative of U is large or (3) the estimate x is biased

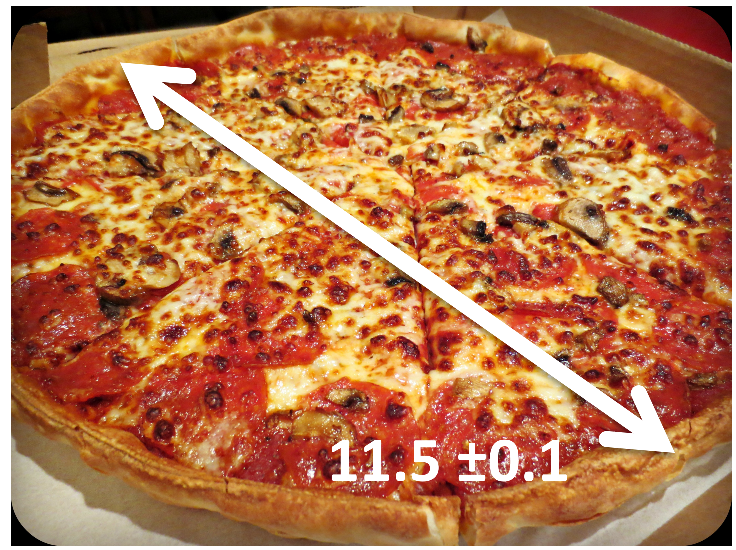
(b) we will operate under the assumption that the bias is negligible for the purposes of this course

Example – Area of a Pizza

We've measured the diameter D of a pizza several times and discovered that the average was 11.5 inches and the sd was 0.1 inches.

Estimate the area of the pizza and

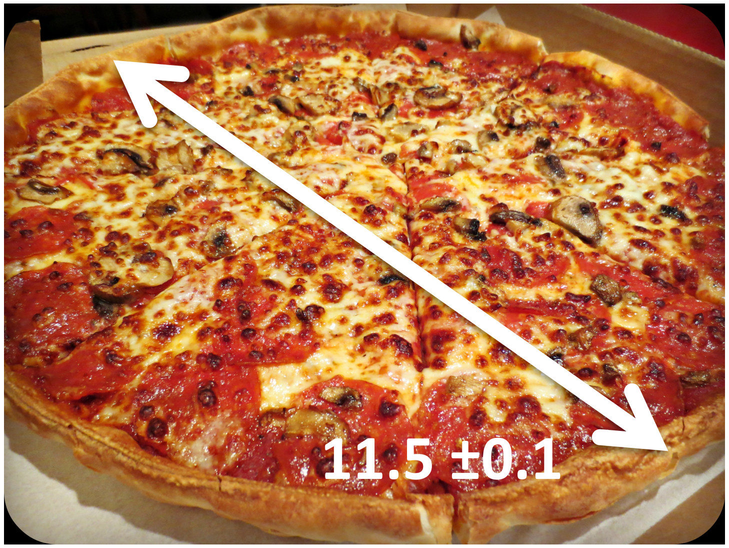
find the uncertainty in that estimate.



Example – Area of a Pizza

We've measured the diameter D of a pizza several times and discovered that the average was 11.5 inches and the sd was 0.1 inches.

Estimate the area of the pizza and find the uncertainty in that estimate.



Estimate: A = 103.87in^2

Uncertainty: $\sigma_A \approx 1.81$ in^2

Uncertainties for (Nonlinear) Functions of Several Measurements

- The propagation of error formula generalizes for multiple **independent** measurements $X_1,\ldots,X_n$ estimated by $x_1,\ldots,x_n$ with (small) uncertainties $\sigma_{X1},\ldots,\sigma_{Xn}$ and a function $U = U(X_1,\ldots,X_n)$:

$$\sigma_U \approx \sqrt{\left(\left.\frac{\partial U}{\partial X_1}\right|_{\vec{X}=\vec{x}}\right)^2 \sigma_{X_1}^2 + \ldots + \left(\left.\frac{\partial U}{\partial X_n}\right|_{\vec{X}=\vec{x}}\right)^2 \sigma_{X_n}^2}$$

$$\text{where } \vec{X} = (X_1,\ldots,X_n) \text{ and } \vec{x} = (x_1,\ldots,x_n)$$

- If the measurements are NOT independent, we can only get an estimate of the upper bound for the uncertainty

$$\sigma_U \leq \left|\left(\left.\frac{\partial U}{\partial X_1}\right|_{\vec{X}=\vec{x}}\right)\right| \sigma_{X_1} + \ldots + \left|\left(\left.\frac{\partial U}{\partial X_n}\right|_{\vec{X}=\vec{x}}\right)\right| \sigma_{X_n}$$

Example 3.18 – Density of a Rock

The mass (m) and volume (V) of a rock are measured repeatedly and found to be ***m* = 674.0 ± 1.0g and V = 261.0 ± 0.1mL**. Assume that *m* and V are independent.

Estimate the density D and find its uncertainty.

What if *m* and V are not independent?

Example 3.18 – Density of a Rock

The mass (m) and volume (V) of a rock are measured repeatedly and found to be ***m* = 674.0 ± 1.0g and V = 261.0 ± 0.1mL**. Assume that *m* and V are independent.

Estimate the density D and find its uncertainty.

2.582 ± 0.004 g/mL

What if *m* and V are not independent?

$\sigma_D \leq 0.0048$ g/mL

Next

- Note: we are not going to cover the topic of relative uncertainties (pg 182 for single measurements and pg 190 for multiple measurements)

- Chapter 4 – Distributions!