Confidence, Proportions and Differences

Mark Muldoon

Departments of Mathematics and

Optometry & Neuroscience

UMIST

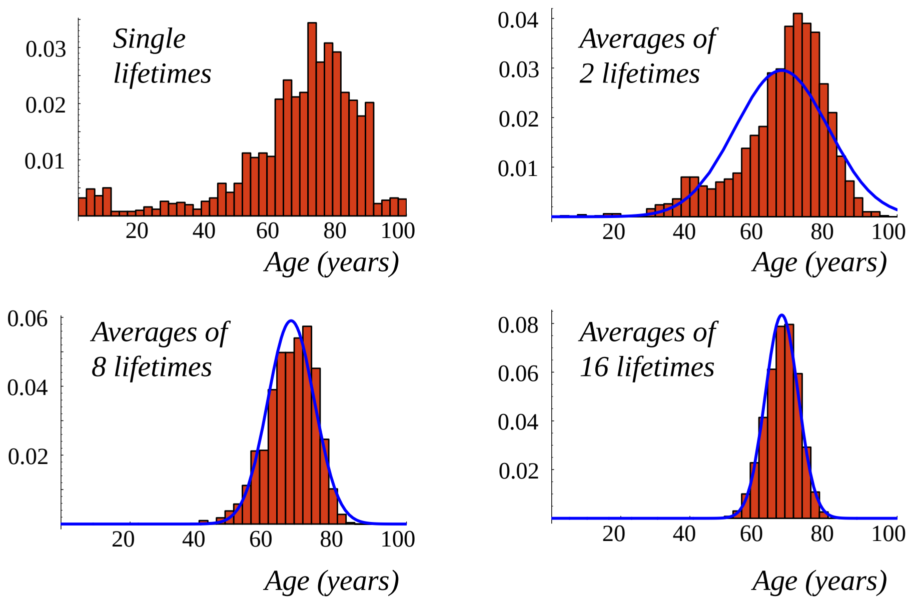
http://www.ma.umist.ac.uk/mrm/Teaching/2P1/

Overview

Today we'll consider further uses of the normal distribution including:

- **Estimating means:** the Central Limit Theorem in action;
- **Proportions:** the Normal approximation to the Binomial distribution and the estimation of proportions;
- **Differences between proportions:** a first look at hypothesis testing.

Estimating population means



Suppose we were trying to estimate the average lifespan of English men—what conclusions could we draw from this figure?

Distributions of averages

Say original distribution had mean μ and standard deviation σ:

- averages over large samples (though still small compared to the whole population) tend to become normally distributed;
- normal distribution obeyed by averages over N values has mean $\mu_N = \mu$ the same as the original distribution, but $\ldots$
- distribution of averages has standard deviation

$$\sigma_N = \frac{\sigma}{\sqrt{N}}.$$

Standard error of the mean

Given a large sample of N values $\{x_1,\, x_2,\, \ldots\; x_N\}$, assume variance of the sample is a good estimate of the variance of the whole population:

$$\sigma^2 \;\approx\; s^2 \;=\; \frac{\sum_{j=1}^{N}(x_j - m)^2}{N-1}.$$

The quantity

$$(s/\sqrt{N}) \approx (\sigma/\sqrt{N})$$

is called the *Standard Error of the Mean*, or *SEM* for short.

Check: SEM

True or false?

The standard error of the mean of a sample:

(a) measures the variability of the observations;

(b) is the accuracy to which each observation is measured;

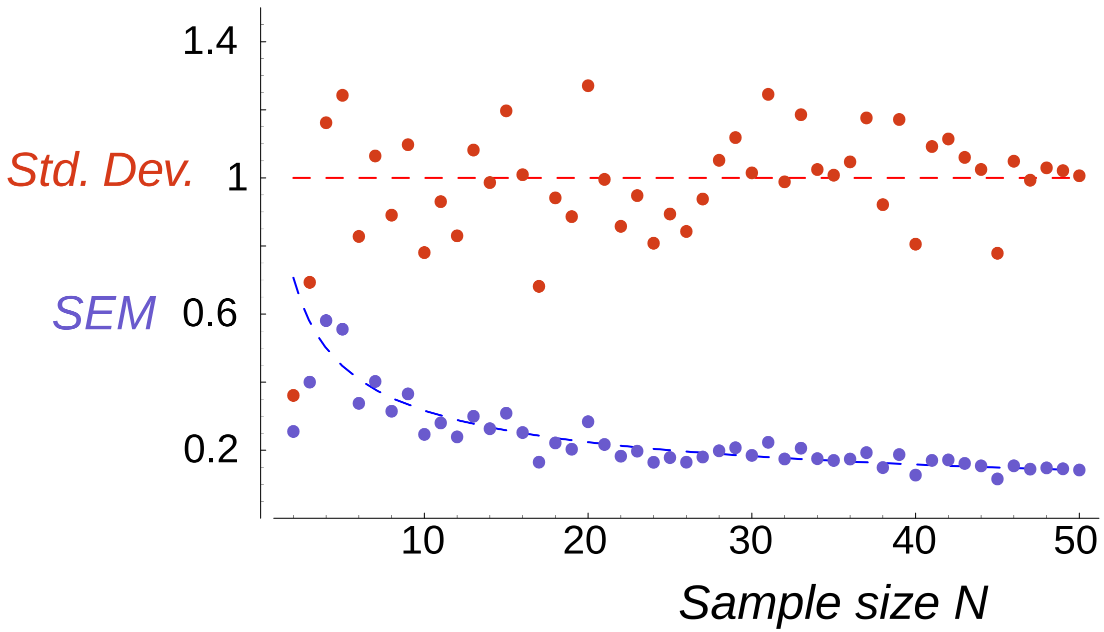
(c) is proportional to the number of observations;

(d) is greater than the estimated standard deviation of the population;

(e) is a measure of how far the sample mean is likely to be from the population mean.

Answer:

SEM &. Std. Deviation: a simulatrion study



The sample standard deviation s and SEM $= s/\sqrt{N}$ for variously sized samples drawn from the standard normal ($\mu = 0$, $\sigma = 1$) distribution. Notice that although s is usually close to σ, the SEM decreases toward zero as the sample size increases.

Estimating the population mean

What can we say about population mean?

Our best estimate is, of course, the sample mean

$$\mu \approx m = \frac{\sum_{j=1}^{N} x_j}{N},$$

but one can say more using z-scores and the Central Limit Theorem.

Probability of error

Q1: What's the probability that the sample mean is differs from the population mean by an amount δ?

A1: Use a z-score based on Central Limit ideas:

$$z = \frac{(m - \mu)}{(\sigma/\sqrt{N})} \approx \frac{\delta}{(s/\sqrt{N})}$$

where

- $(m - \mu)$ is the difference between the sample mean and the true mean of the (approximate) normal distribution obeyed by the sample mean;
- $(\sigma/\sqrt{N}) \approx (s/\sqrt{N})$ is the standard deviation of the (approximate) normal obeyed by the sample mean.

How wrong could we be?

Q2: What's an implausibly large value of δ?

A2: One that corresponds to an implausibly large z-score. For example, $P(z \geq 1.96) \approx 0.025$ which means

$$P\left(\left(\frac{\delta}{(\sigma/\sqrt{N})}\right) \geq 1.96\right) \approx 0.025$$

or, using our assumption $\sigma \approx s$,

$$P\left(\delta \geq 1.96 \times \text{SEM}\right) \approx 0.025.$$

Not so wrong (probably)

And so . . .

$$\begin{aligned} P\left(|\delta| \geq 1.96 \times \text{SEM}\right) &= \\ &= P\left(\left(\delta \leq -1.96 \times \text{SEM}\right) \textit{ or } \left(\delta \geq 1.96 \times \text{SEM}\right)\right) \\ &= P\left(\delta \leq -1.96 \times \text{SEM}\right) \\ &\quad +P\left(\delta \geq 1.96 \times \text{SEM}\right) \\ &\approx 0.025 + 0.025 = 0.05 \end{aligned}$$

There's only a 5% chance of seeing such a big δ.

Numerical confidence

Or, a bit more usefully

$$P\left(-1.96 \times \text{SEM} \leq (m - \mu) \leq 1.96 \times \text{SEM}\right) \approx 0.95.$$

Or, most usefully of all

$$P\left((m - 1.96 \times \text{SEM}) \leq \mu \leq (m + 1.96 \times \text{SEM})\right) \approx 0.95$$

We have constructed an interval that (probably) contains the population mean! This is an example of a *95% confidence interval*.

Other levels of confidence

By choosing different cutoffs for z one can construct intervals that contain the mean with various levels of confidence. The recipe is

(1) Choose a confidence level C ($C = 0.95$ for 95%) and define $\alpha = 1 - C$.

(2) Use the standard normal table to find that z-score, $z_{\alpha/2}$, such that

$$P(\, z \geq z_{\alpha/2}\,) = (\alpha/2).$$

(3) If your sample has mean m the desired confidence interval for the population mean μ is:

$$m - z_{\alpha/2} \times \text{SEM} \leq \mu \leq m + z_{\alpha/2} \times \text{SEM}$$

Example: at a national conference

A speaker mentions that the UK average annual salary of new graduate engineers is £18,452. Suppose your circle of local colleagues includes 37 starting engineers whose mean salary is $m =$ £17,283 with a standard deviation of $s =$ £2,210.

Q: Are you and your colleagues underpaid?

Hint: *Construct, say, a 90% confidence interval for the mean salary in your region. Does it include the national average?*

Proportions and the normal

In N independent trials, each with a probability p of "success", the distribution of the *number* of successes is approximately normal with

$$\mu = Np \qquad \text{and} \qquad \sigma = \sqrt{p(1-p)N}.$$

The distribution of the *proportion* of successes is thus approximately normal with

$$\mu = p \qquad \text{and} \qquad \sigma = \sqrt{p(1-p)/N}.$$

These approximations are applicable when both pN and $(1-p)N$ exceed 5.

Testing proportions (or the lecturer's sickly, ancient car)

You are studying a fault known to occur 88 old Rovers per every 1000 in the UK. You arrange to screen samples of 250 cars at various places throughout the country.

(a) How would you decide whether an area had an unusual number of malfunctioning cars?

(b) How would you decide if an area had an usually *large* number of malfunctioning cars?

Detecting unusual places

This is essentially a job for z-scores:

(a) Choose a confidence level, say $C = 0.95$

(b) Probability a car has the fault is $p = 0.088$ so distrib, of number of cases is approximately normal with

$$\begin{aligned} \mu &= 250 \times 0.088 = 22 \\ \text{and} \quad \sigma &= \sqrt{250 \times 0.088 \times (1 - 0.088)} \\ &\approx 4.48 \end{aligned}$$

(c) Construct suitable confidence interval and classify as extreme those places where the number of cases falls outside:

$$14 \leq \text{ Number of cases in 250 } \leq 30.$$

Classifying unusually large numbers

The reasoning is similar, but now we are only concerned with deviations to one side.

(1) Choose a confidence level, say $C = 0.95$ and, as usually, define $\alpha = 1 - C$.

(2) Use z-score in reverse to choose a *critical value* n_α for the number of faulty Rovers so that

$$\frac{n_\alpha - \mu}{\sigma} = z_\alpha$$

or

$$n_\alpha \;=\; \mu + z_\alpha \times \sigma \;\approx\; 22 + 1.64 \times 4.48 \;\approx\; 29.35$$

So one would use 30 as a conservative cutoff.

Hypothesis testing

The previous problem was an example of *Hypothesis Testing:*

- Choose some default explanation for the data you are about to collect (frequency of rejected product at this plant is the same as the company average, experimental coating has no consequence . . .) and call it the *null hypothesis*.
- Choose some other explanation (local wastage is lower than company average, new coating inhibits corrosion) that you would like to test for: call this the *alternative hypothesis*.
- Compute some test statistic for the actual data.
- Work out the probability of getting the observed value of the test statistic if the null hypothesis were true.
- *Reject* the null hypothesis if the observed value is too unlikely.

Possibilities for error

One may come to the wrong conclusion in a hypothesis test . . .

		Our decision:	
		Accept	*Reject*
Null is:	*True*	Correct Decision $1-\alpha$	Type I Error α
	False	Type II Error β	Correct Decision $(1-\beta)$ a.k.a. the *power*

Undeserved rejection: α

What is the risk a Type I Error: rejecting the null hypothesis when it is really true?

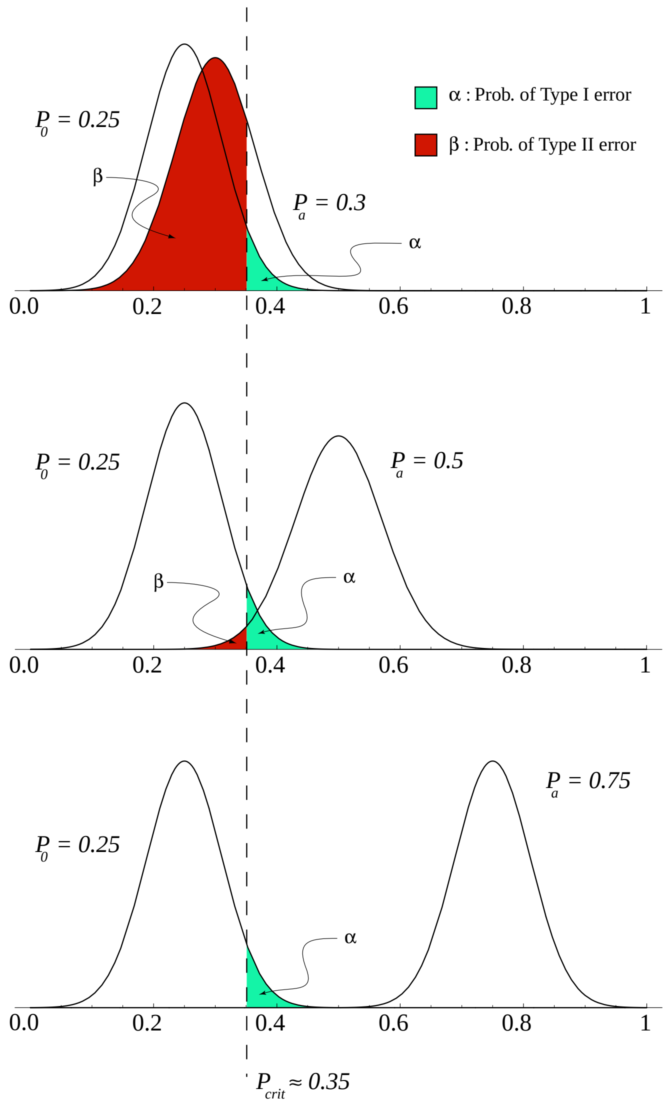
- It's α, which is entirely under the analyst's control;
- it does not depend on the sample size, only on the desired level of confidence;
- it depends only on the null hypothesis and properties of the distribution of the test statistic under the null.

Unfounded acceptance: β

What is risk of a Type II error: accepting the null hypothesis when it is false?

- it is called β—the quantity $(1 - \beta)$ is called the “power” of the test;
- it depends on “how false” the null is, which depends on the sample size;
- this not under the data analyst’s control.

Type II errors for frequencies



Normal approximations to the distributions of results for frequency measurements

Each panel shows two distributions of results, those expected if the null hypothesis is true

$$P_a = P_0 = 0.25$$

and those expected from some other local frequency: P_a = 0.3 (top), 0.5 (middle) and 0.75 (bottom).

Probabilities of errors

This table summarizes the shaded areas (probabilities of error) shown on the previous slide.

				Confidence Limit
Panel	P_a	β	α	$(1 - \alpha)$
Top	0.3	78.3%	5%	95%
Middle	0.5	1.7%	5%	95%
Bottom	0.75	3.5×10^{-9}%	5%	95%