AMS 7 - Lecture 5·8·18

THIS TIME → inference for μ

NEXT TIME → Inference for P

Read → LN pg. 137-160 today → pg. 137 →

DD ch. 1-11 (B)

DD ch. 1-3 (A)

4.1 ▷ Confidence Intervals

*case study (marine biology)

- Question 1:

  Is the difference between 25.0°C ($\bar{y}$) & 24.3°C (theoretical mean) practically significant?

  ↳ Large in biological terms

- Answer 1:

  (Best) Consult an expert on intertidal crabs

  (helpful) $\frac{25.0°C - 24.3°C}{24.3°C} = 2.9\%$

  *Is a **2.9%** diff. between data & theory big enough to matter?

→ Rough rule of thumb:

relative differences of 5% or more are often, but not always, Practsig; be Practsig if they accumulate over time

Intertidal Crabs

Population | Sample | Imaginary data

(All crabs similar to those in sample in all relavent ways) | (The observed crabs)

$N = ?$ — Has to be big

like SRS → $\doteq$ IID

e. temp

$\begin{bmatrix} 25.8 \\ \vdots \\ 25.4 \end{bmatrix}$ $n = 25$

$\mu = ?$

$\sigma = ?$

mean $\bar{y} = 25.0°C$

SD $s = 1.34°C$

*To specify population, Answer what is the broadest scope of valid generalizability outward from the sample data set?

Hypokalemia

Population $\xrightarrow{IID}$ Sample | **Probability** (easier)

Whole known — Part unknown

General — Specific (particular)

μ = known — $\bar{Y}$ = unknown

(deductive reasoning) = (deduction)

CRABS

Population $\xrightarrow{IID}$ Sample | **Statistics**

Whole unknown — Part known

General ← Particular

μ = unknown (?) — $\bar{y}$ = known

(inductive reasoning) = (induction)
= (statistical inference)

The Inferential Summary

	Crabs	
Pop.	unknown pop. quantity of main interest	μ = Pop. mean internal temp after equil. to 24.3°C
Sample	estimate of μ	$\bar{y}$ = 25.0°C
Imagin. data set	Give-or-take for $\bar{y}$ as est. of μ	$\hat{SE}(\bar{y}) = \frac{s}{\sqrt{n}} \doteq 0.27°C$
	Interval 95% Confidence	$\bar{y} \, I \, (2.064)(0.27°C) = (24.44°C, 25.56°C)$ $= (25.0°C - 0.56°C, 25.0°C + 0.56°C)$

Population | Sample | Imaginary data set

Actual like IID $\rightarrow$ [ ] $n = 25$

SD $s = 1.34°C$

mean $\bar{y} = 25.0°C$

mean $\mu = ?$

SD $\sigma = ?$

Pop. Histogram ?

R-22

*

Hyp. IID [ ] $n = 25$

mean $\bar{y} = ?$ (ex. 25.2°C)

$\rightarrow$ more Hypothetical Samples

All possible values of $\bar{y}$.

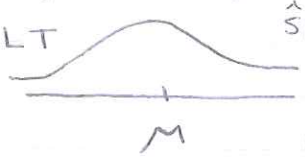
$$\begin{bmatrix} 25.0°C \\ 25.2°C \\ \vdots \end{bmatrix} \begin{matrix} \bar{y}_1 \\ \\ M \rightarrow \infty \end{matrix}$$

long run mean (EV of $\bar{y}$) $= \mu$

long run SD (SE of $\bar{y}$) $= \cancel{\sigma} / \sqrt{n} = s/\sqrt{n} = 0.27°C$

long run histogram of $\bar{y}$

CLT $\hat{SE} = 0.27°C$



give or take for $\bar{y}$ of 1 as an estimate of μ

CLT? Part 3

* expected value = EV of $\bar{y} = E_{IID}(\bar{y}) = \mu$

* standard error of $\bar{y}$ = SE of $\bar{y}$ = $SE_{IID}(\bar{y}) = \frac{\sigma}{\sqrt{n}}$

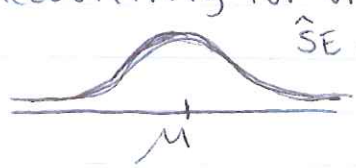
$\hookrightarrow \hat{SE}(\bar{y}) = \frac{s}{\sqrt{n}} = \frac{1.34°C}{\sqrt{25}} = 0.27°C$

>>> this is the estimated SE of $\bar{y}$

$\rightarrow$ State of Statistical Inference as of 1907

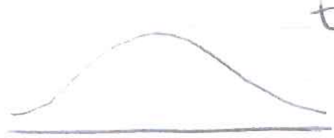
- William Gosset (1908)
  - $\hookrightarrow$ worked for Guiness Brewery
    - *Hops, yeast, barley, water = Beer!
    - *He noticed a flaw with the model long run histogram of $\bar{y}$ (t curve) accounting for uncertainty in σ

$\hat{SE}$ 0.27°C



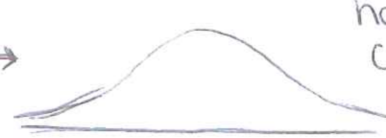
$\underbrace{t_{n-1}}_{\text{degrees of freedom}}$ curve

L-142
t Table



t_{γ} curve
└ nu "new"

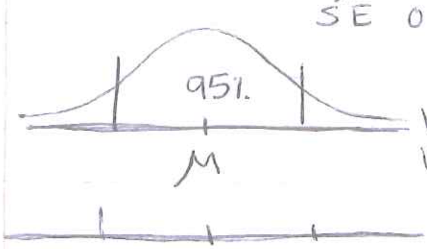
→



$n = 25$
$n - 1 = 24$

t_{24} curve

→ long run histogram of $\bar{y}$, accounting for uncertainty in σ
$\hat{SE}$ 0.27°C



$M + 2.064\ \hat{SE} = M + (2.064)(0.27°C)$

$P_F\,(M - 2.064\,\hat{SE} \le \bar{y} \le M + 2.064\,\hat{SE}) \doteq 95\%$
↑
relative
frequency
(probability)

*Neyman's confidence trick
$P_F\,(\bar{y} - 2.064\,\hat{SE} \le M \le \bar{y} + 2.064\,\hat{SE}) \doteq 95\%$

→ lets use $\bar{y} \pm \underbrace{2.064}_{t_{n-1}\ 0.95}\ \hat{SE}$ as a 95% $\overbrace{\text{Confidence Interval}}^{CI}$ (for M)

24.3°C
↓

95% CI for M

*at 95% level of confidence, the theory value of 24.3°C is <u>not</u> supported by the data; The diff. between $\bar{y}$ (25.0°C) & M_0 (24.3°C) ....

is statistically significant (statsig).

μ_0 = theory value of $\mu = 24.3°C$