this time: hypothesis & significance testing

next time: sample size determination

Ashley Tai
AMS 7
15 May 18

$\boxed{\text{HW 3 due on canvas by 11:59 pm on Sun 27 May}}$

read:
LN pp. L-174 → L-185   today: LN pp. L-162 →
L-139 Intertidal Crabs

Neyman's logic: try null on for size; see if discrepancy b/t (how data came out) vs (how data should have come out if null true) is large.

- if yes, favor alternative (reject null)
- if not, favor null (fail reject null)

null hypothesis
$\mu_0 = 24.3°C$
alternative hyp
$\mu \neq 24.3°C$

↳ pretend pop. mean $\mu = 24.3°C = \mu_0$

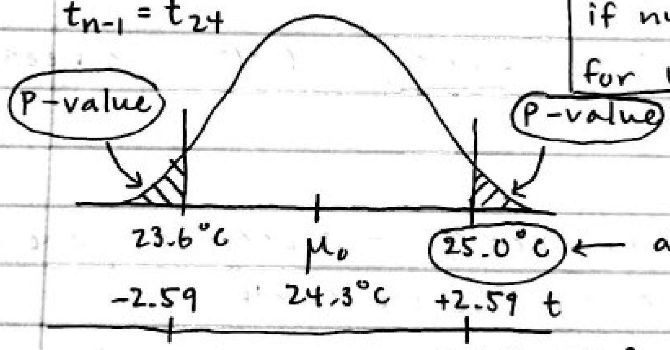
thus:

- long run mean = EV of $\bar{y}$ = 24.3°C
- long run histogram of $\bar{y}$ if null true

$\hat{SE} = 0.27$

$\boxed{\text{long run hist. of } \bar{y} \text{ if null true, accounting for uncertainty in } \sigma}$

$t_{n-1} = t_{24}$



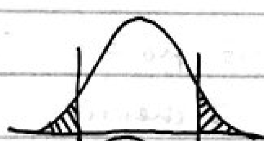
$$\text{t statistic} = \frac{25.0 - 24.3}{0.27} = \frac{\bar{y} - \mu_0}{\hat{SE}(\bar{y})} = \frac{0.7\cancel{°}}{0.27\cancel{°}} = \underline{\underline{2.59}} = t$$

$$= \text{t ratio} = \frac{\text{signal}}{\text{noise}}$$

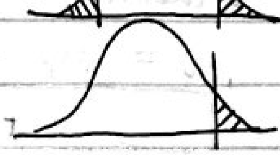
P-value: the chance, if null true, of getting data as extreme as, or more extreme than what I got

— measure of surprise

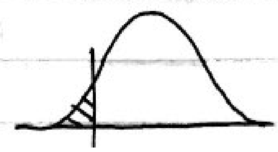
$\left(\begin{matrix}\mu \neq \mu_0 \\ \text{alt}_1\end{matrix}\right)$: two-sided alternative
2-tailed P-value



$\left(\begin{matrix}\mu > \mu_0 \\ \text{alt}_2\end{matrix}\right)$: one-sided alternative
1-tailed P-value



$\left(\begin{matrix}\mu < \mu_0 \\ \text{alt}_3\end{matrix}\right)$: one sided alternative
1-tailed P-value



★ if p-value is small, favor alternative & reject null

★ if p-value is big, favor null

$\boxed{Q1}$: How small?

$\boxed{A}$: People are lazy, so:

* if $p \leq 5\%$ → statistically significant
* if $p \leq 1\%$ → highly statsig

} reject null

here, $p = 1.6\%$, so the diff. b/t $25.0°C$ ($\bar{y}$) & $24.3°C$ (μ_0) is statsig, but not highly statsig

* most of the time in science, 2-sided alt. is more appropriate

CIs are better than p-values

- can't tell if $\bar{y}$ is above or below μ_0
- no way to tell how big signal ($\bar{y} - \mu_0$) is & how big noise ($s/\sqrt{n}$) is
- ($\bar{y} - \mu_0$) is needed to judge practsig & ($s/\sqrt{n}$) is needed to judge statsig

★ when diff is statsig & not practsig, it is because there is too much data (n too large & SE too small)

★ when diff is practsig & not statsig, it is because there is not enough data (n too small & SE too large)

Sample size determination

1) CI approach

$$\bar{y} \pm \left(t_{n-1}^{0.95}\right) \frac{s}{\sqrt{n}}$$

• if $\mu_0 = 32$ falls just outside 95% CI, CI will discriminate b/t 2 theories

$$\mu_0 = \mu_A + \left(t_{n-1}^{0.95}\right) \frac{s}{\sqrt{n}} \rightarrow \text{solve for } n$$

(1): 1-tailed test

(2): 2-tailed test

$$\boxed{n = \frac{\left(t_{n-1}^{(1-\alpha)(2)}\right)^2 s^2}{(\mu_0 - \mu_A)^2}}$$

$s \uparrow$	$n \uparrow$
$\lvert\mu_0 - \mu_A\rvert \uparrow$	$n \downarrow$
$\alpha \uparrow$	$n \uparrow$

* to solve iteratively :
start on right hand side (rhs) with $n = \infty$ ($t \doteq z$ (normal curve)), solve for n, look up new t # with this n & put it in rhs, solve again, & repeat as needed

2) significance / hypothesis testing approach

H_0 (null) : $\mu = \mu_0$ (theory 1) $(\mu_0 = 32)$

H_A (alt) : $\begin{cases} \mu \neq \mu_0 & \text{(2-sided alt)} \quad \text{(2-tailed test)} \\ \begin{cases} \mu > \mu_0 \\ \mu < \mu_0 \end{cases} & \text{(1-sided alt)} \quad \text{(1-tailed test)} \end{cases}$

type I error : false rejection of null

want this small

$P(\text{reject } H_0 \mid H_0 \text{ true}) = P(\text{type I error}) = \alpha$ (significance level)

type II error : false acceptance of null

$P(\text{don't reject } H_0 \mid H_0 \text{ false}) = P(\text{type II error}) = \beta$ (Beta)

Power of the test = $(1 - \beta)$ ← want this big

* in order to make P(type I error) < α & P(type II error) < β, n needs to be at least:

$$\boxed{n = \frac{\left[t_{n-1}^{(1-\alpha)(?)} + t_{n-1}^{(1-\beta)(1)}\right]^2 \cdot s^2}{(\mu_0 - \mu_A)^2}}$$

2 - Sample Inference

- case 1 : paired comparisons
- case 2 : analysis of 2 independent samples