AR
pg 1/4

Feb. 17, 2009

Hypothesis testing (continued)

· How do we measure discrepancy? t test

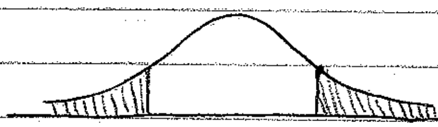
$$\frac{s}{\sqrt{n}} \rightarrow \frac{\bar{y} - \mu_0}{\left(\text{SE of } \bar{y} \text{ if null were true}\right)} = \frac{25°C - 24.3°C}{.27°C} = \frac{.7°C}{.27°C} = \frac{\text{"signal"}}{\text{"noise"}} = \boxed{2.59 = t}$$

"t statistic"

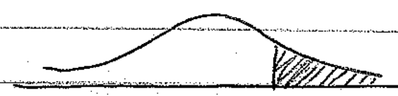
· chance, if null true, of getting data as extreme as, or more extreme than, what we got = numerical suprise value = P value
- if big P, favor null ; little P, favor alternative

Alternative Hypotheses

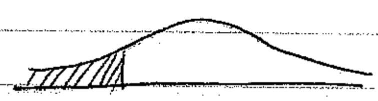
$\boxed{1}$ $\mu \neq \mu_0$: two sided alternative (2 ~~side~~ tailed test)
(alt$_1$) 2 tailed p value



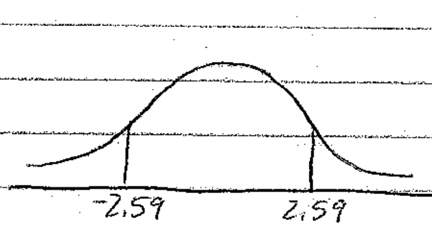
$\boxed{2}$ $\mu > \mu_0$: one sided alternative (1 tailed p value)
(alt$_2$)



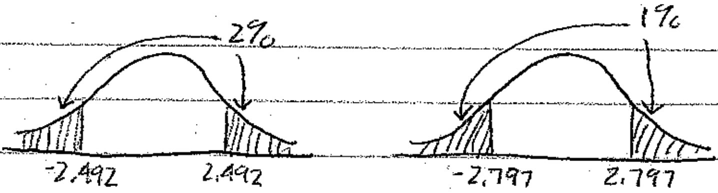
$\boxed{3}$ $\mu < \mu_0$: one sided alternative (1 tailed p value)
(alt$_3$)



$\longrightarrow$ t_{24} curve



From t table:



Feb. 17, 2009

- 2.59 falls between 2.492 & 2.797, so $1\% < P < 2\%$

- How small is small enough for a p-value in order to reject null and favor the alternative?
  - if $P \leq 5\%$, result is statistically significant
  - if $P \leq 1\%$, result is highly statistically significant

In our example, $1\% \leq P \leq 2\%$, so the difference between 25°C (y) and 24.3°C (μ_o) is statistically significant.

* Math Fact: When hypothesis testing is done with 2-sided alternative, its conclusion is identical to that of CI approach.

- Pitfalls of hypothesis testing:
  - much more complicated than CI
  - rigid adherence to $p \leq 5\%$ is silly
    - some journals are so rigid about stat. sig. that they will only accept a paper if $P \leq 5\%$. Suppose you do a test and get a 2-tailed p-value of 8%, can't publish. But, to get a smaller p-value, you can just pretend that the real alternative was 1-sided so the p-value $= 8\%/2 = 4\%$. Suddenly it's okay to publish.

- Why CI's are better than p-values

ex: crab data

- with hypothesis testing, 2-tailed $p = 1.6\%$

  If all you know is this, you can't tell if $\bar{y}$ came out above or below μ_o.

Feb. 17, 2009

- Practical and Statistical Significance
  - need to know $(\bar{y} - \mu_0)$ to judge practical significance and need to know $(s/\sqrt{n})$ to judge statistical significance.
  - Can we judge practical significance from CI? Yes
  - Can we judge statistical significance from CI? Yes

Ex: New drug to reduce blood pressure
take BP before and after

Person #	Before	After	Difference (A-B)
1	177	159	-18
2	155	157	+2
⋮	⋮	⋮	⋮
n	201	201	0

mean $\bar{y} = -1$ mm Hg
SD $s = 10$ mm Hg

- Difference of 1 mmHg is not practically significant
- Is it statistically significant?

95% CI : $-1 \pm .6 \leftarrow \bar{y} \pm t_{999}^{.95}\left(s/\sqrt{n}\right)$
$-1 \pm 1.96\left(10/\sqrt{1000}\right)$

-1.6   -1   -.4

* not in 95% CI, so yes, statistically significant

- Why is it stat sig but not prac sig?
  Too much data
- Can also be prac sig without being stat sig.
  Caused by too little data.

- Sample Size Determination:

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

Feb. 17, 2009

* needs to be solved iteratively
- start on right hand side with $n = \infty$, solve for n, look up new t with this n and put it in right hand side, solve again, and repeat as needed.

Ex: $s = 1.8$

$|\mu_0 - \mu_A| = (32 - 31.5) = .5$

$\alpha = .05$ + 95% CI

1) $n = \dfrac{(1.961)^2(1.8)^2}{(.5)^2} = 49.8 = 50$ (always round up with n)

2) try $n = 50$: $t_{n-1}^{.95(2)} = t_{49}^{.95(2)} = 2.010$

$$n = \frac{(2.010)^2(1.8)^2}{(.5)^2} = 52.4 = 53$$

3) try $n = 53$: $t_{n-1}^{.95(2)} = t_{52}^{.95(2)} = 2.007$

$$n = \frac{(2.007)^2(1.8)^2}{(.5)^2} = 53.$$ done

$\boxed{n = 53}$