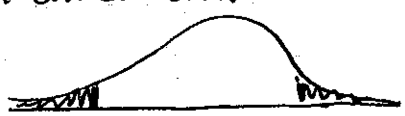
(1) Lecture #12 - Hypothesis Tests Feb 17, 2009

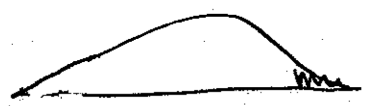
- If P-value is small, favor alternative; if p is big, favor the null

(alt 1, $\mu \neq \mu_0$): two-sided alternative (2-tailed test)



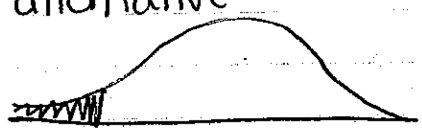
(2-tailed P value)

(alt 2, $\mu > \mu_0$): one-sided alternative (1-tailed test)



(1-tailed P value)

(alt 3, $\mu < \mu_0$): one-sided alternative (1-tailed test)



(1-tailed P value)

t_{24} curve

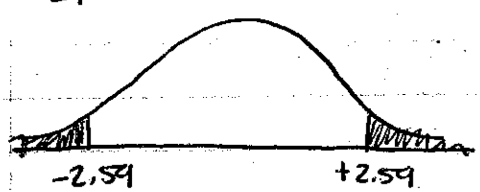
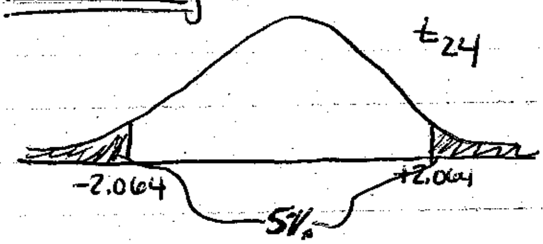
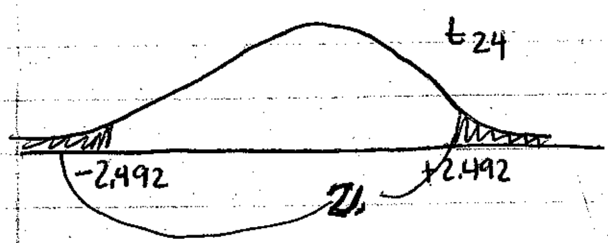
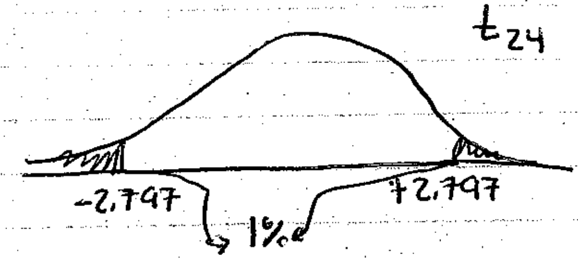


Table Says:







here, $1\% \leq p \leq 2\%$ (closer)

(2) Lecture #12 — 17 Feb, 2009

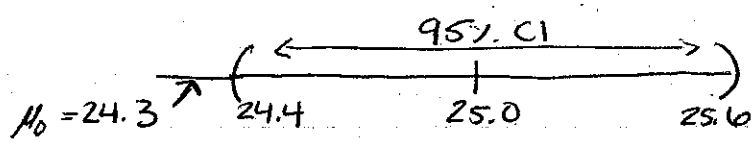
Q: How small is small enough for a p-value in order to "reject null" & favor alternative?

A: This is difficult to figure out & people are lazy so people appeal to conventional values instead:

if $p \le 5\%$ result is statistically significant

if $p \le 1\%$ result is highly statsig

Here $p = 1.6\%$ (from JMP), so the difference between 25.0°C ( ) & 24.3°C (μ_0) is statsig but not highly statsig.



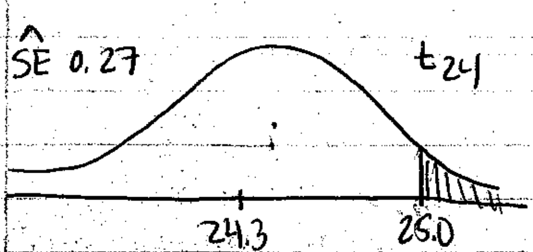
CI approach: see if μ_0 is in 95% interval; if not, difference is statsig.

Math fact: when hyp. testing is done with 2-sided alternative, its conclusion is identical to that of CI approach

CI is better: simple & gives more information

Pitfalls of hyp testing

$\begin{pmatrix}\text{2 sided alt.}\\ \mu \ne \mu_0\end{pmatrix} \longleftrightarrow \begin{pmatrix}\text{2-tailed t test =}\\ \text{2 tailed P value (1.6\%)}\end{pmatrix}$



1-tailed $P = \frac{1.6\%}{2} = 0.8\%$

$\begin{pmatrix}\text{1 sided}\\ \text{alt}\end{pmatrix} \longleftrightarrow \begin{pmatrix}\text{1-tailed}\\ \text{test}\end{pmatrix}$

(3) <u>Lecture #12</u>

<u>cont</u>...

$H_0: \mu = \mu_0$ 1 sided alternative

$H_A: \mu > \mu_0$

* Frequent use of hyp test:
  - null: my theory wrong
  - alt: my theory right

Here, I want to reject <u>null</u>, I want a small p-value; some jurnals are so rigid that they only accept paper if $p \leq 5\%$. If I do my test with my data and I get a 2-tailed p value of 8%, I can't publish. So, a quick way to get a smaller p-value: pretend that real alt. was 1 sided; then 1-tailed p-value = $8\% \div 2 = 4\%$ and suddenly it's okay to publish. (weakens hypothesis theory)

$\boxed{\text{Problems}}$

(a) rigid adherence to $P \leq 5\%$ is "silly"

(b) If you would be convinced against null by 1-tailed $p = 4\%$, you should be equally convinced by 2-tailed $p = 8\%$

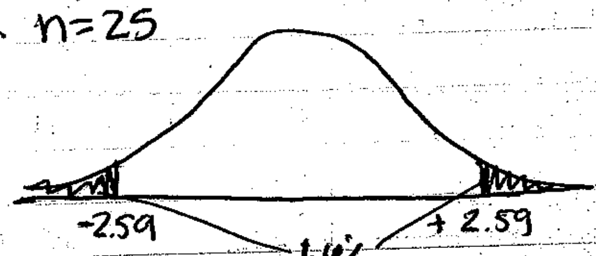
most of the time in science, 2-sided altn. is more appropriate

* CI's also better than p-values: crab data:

2-tailed $P = 1.6\%$ $n = 25$

null: $\mu = 24.3°C$

alt: $\mu \neq 24.3°C$



† If this is all you know you cannot reconstruct $\bar{y}$

(4) Lecture # 12

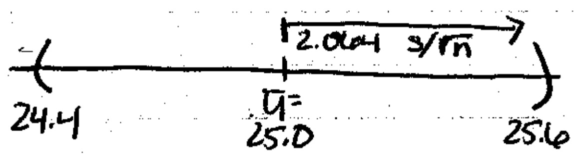
<u>Problems</u> cont... (with hypothesis tests)

$$t = \pm 2.59 = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{\bar{y} - 24.3}{s/\sqrt{25}}$$

$= \frac{\text{signal}}{\text{noise}}$; no way to tell how big signal $(\bar{y} - \mu_0)$ is or how big noise is $(s/\sqrt{n})$

- It mushes two parts together that we need separatley. Need to know $(\bar{y} - \mu_0)$ to judge practical significance & $s/\sqrt{n}$ for statsig

<u>Instead</u>: 95% CI (24.4, 25.6) n = 25
$= \bar{y} \pm t_{n-1}^{.95} \, s/\sqrt{n}$



(diff. between $\bar{y}$ & $\mu = 24.3$)

Here we can work out the variables of the equations and know if the data is practically significant and/or statistically significant.

"<u>Plus</u> you can test by looking to see if the value of interest is in the interval or not.

"<u>Plus Plus</u> - if a new theoretical value is brought up for μ_0 can see if new μ_0 is or is not in the 95% CI.

(5) Lecture #12

+ Statsig ≠ practsig

ex: New drug to reduce blood pressure of hypertensive patients
"Repeated Measures Design
[Before (B), After(A)] for each person

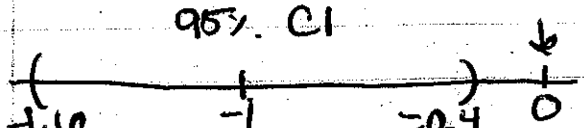
Person #	B	A	Diff A-B
1	177	159	−18
2	155	157	+2
⋮	⋮	⋮	⋮
n	201	201	0

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

−the difference of 1 mmHg is not practically significant (clinically, medically)

$\boxed{Q}$ Statsig? devil's advocate null: $\mu = 0 \leftarrow \mu_0$
popmean diff(A−B)

95% CI:
−1 ± .6



$\bar{y} \pm t_{999}^{.95} \frac{s}{\sqrt{n}}$

$-1 \pm 1.96 \underbrace{\frac{10}{\sqrt{1000}}}_{.32}$

here: t=z because n is so large

0 is not in the 95% CI so the difference is statsig.

17 Feb 2009

(6) <u>Lecture # 12</u>

$\boxed{Q}$ Why statsig & not practsig?

$\boxed{A}$ Too much data, n to large (SE too small)

$\boxed{Q}$ Can we have practsig but not statsig?

$\boxed{A}$ yes

$\boxed{Q}$ Is this difference practsig?

$D = A - B$

$[\ ]$ $n = 10$

mean $\bar{y} = -10$ mmHg

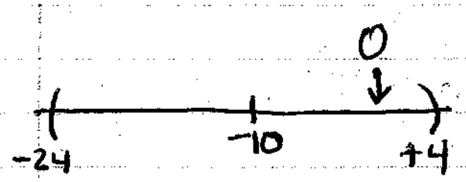
SD $s = 20$ mmHg

$\boxed{A}$ Yes, a 10mmHg decline would matter medically over an extended time period

$\boxed{Q}$ Statsig? $\boxed{A}$ $\bar{y} \pm (t_9^{.95}) \frac{s}{\sqrt{n}}$

$(-10) \pm 2.262 \left(\frac{20}{\sqrt{10}}\right)$

$= (-10 \pm 14)$ mmHg



0 is in the 95% CI so this difference is <u>not</u> statsig

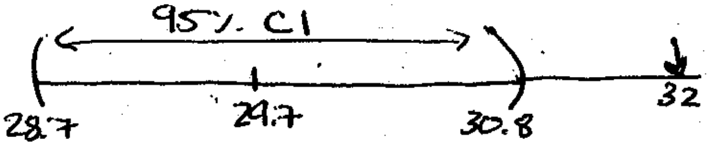
$\boxed{Q}$ Why? $\boxed{A}$ too little data

<u>Solution</u>: choose (n) so that statsig & practsig

(7) Lecture #12 — 17 Feb, 2009

Sample Size Determination (see Discussion Sections #2)

ex:



CI Approach

We do not agree with the $\mu = 32$ theory

μ = pop. mean calcium concentration

null H_0 : $\mu = 32 = \mu_0$ (theory 1)
alt H_A : $\mu = 31.5 = \mu_A$ (theory 2)

Q How big should n be to reliably discriminate between theories 1 & 2?

plan: take (n) obs. & build 95% CI:

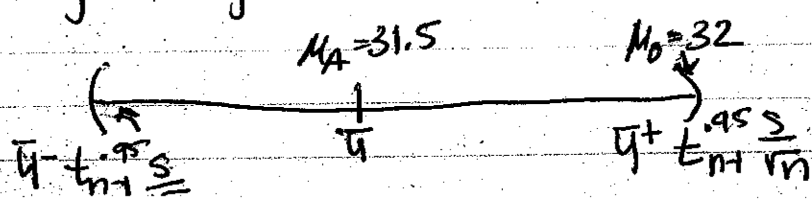
$$\bar{y} \pm \boxed{t_{n-1}^{0.95}} \frac{s}{\sqrt{n}}$$

↳ the place along the t_{n-1} curve with 0.95 in the middle, 0.05 in the 2 tails combined

95% CI = $100(1-\alpha)\%$ ($\alpha = 0.05$)

Suppose Past Experience makes you think (s) will come out around 1.8.

If theory 2 is right:



(8) Lecture #12 17 Feb, 2009

CI will discriminate between the 2 theories if $\mu_0 = 32$ falls just outside the 95% CI:

$$\underset{(32)}{\mu_0} = \underset{(31.5)}{\mu_A} + t_{n-1}^{0.95} \frac{s}{\sqrt{n}},$$ solve for n & get

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

pg R-27 equation # 8

Too much data = too spread out to see importance
To little data = too close to see importance

$s \uparrow$ (noise level of data $\uparrow$) $n \uparrow$ ✓
$|\mu_0 - \mu_A| \uparrow$ (theories are easier to tell apart) $n \downarrow$ ✓
$\alpha \downarrow$ (confidence level of CI $\uparrow$) $n \uparrow$ ✓

* Start on right hand side (rhs) with $n = \infty$ ($t = z$ (normal curve) 1.96), solve for n, look up new t # with this n and put it in rhs, solve again, repeat as needed (but usually only 2x required)

ex. $s = 1.8$ mmol/kg $|\mu_0 - \mu_A| = |32 - 31.5| = 0.5$ mmol/kg

$\alpha = .05 + 95\%$ CI

① $n = \frac{(1.96)^2 (1.8)^2}{(0.5)^2} = 49.8 = 50$ (always round up with n)

② try n=50 $t_{n-1}^{0.95} = t_{49}^{0.95(2)} = 2.010$ (estimate from t-table between t_{45} & t_{50})

$n = \frac{(2.010)^2 (1.8)^2}{(0.5)^2} = 52.4 = 53$

(9) Lecture #12 17 Feb, 2009

③ try n=53 $t_{52}^{0.95(2)} = 2.007$ (estimate from t-table)

$$n = \frac{(2.007)^2(1.8)^2}{(0.5)^2} = 53$$ (done) got same (n) again.

Significance Hypothesis Testing Approach

(null)
$H_0: \mu = \mu_0$ (theory 1) ($\mu_0 = 32$)
H_A: (alt) $\mu \neq \mu_0$ (2 sided alt) (2 tailed test)
$\mu > \mu_0$ / $\mu < \mu_0$ (1 sided alt) (1 tailed test)

(Like credit-card screenings)

test says	truth: H_0 false	truth: H_0 true
reject H_0	good	type I error (α)
don't reject H_0	type II error (β)	good

* Type I error: false rejection of null
* Type II error: false acceptance of null

Nyman & E Pearson:

IF you get your random sample of (n) obs. over and over again, sometimes by chance when H_0 true you would (without meaning to) fall into type I error box; Then times when H_0 is false you might fall into type II error box without meaning to; want both of these error probabilities to be small!