Maya Sertic
2/17/09

HYPOTHESIS TEST (Continued)

We must compare $\bar{Y}$ & μ_0 with respect to the noisiness (uncertainty) of $\bar{Y}$.

$$\left(* = \bar{Y} = 25.0^\circ C\right) \text{ vs. } \left(** = \mu_0 = 24.3^\circ C\right)$$

We assume that the null is true ($\mu = 24.3 = \mu_0$) and do a T-Test:

T-Test: $$\frac{\bar{Y} - \mu_0}{\left(SE_{IID}(\bar{Y}) \text{ if null is true}\right)} = \frac{\bar{Y} - \mu_0}{(s/\sqrt{n})}$$

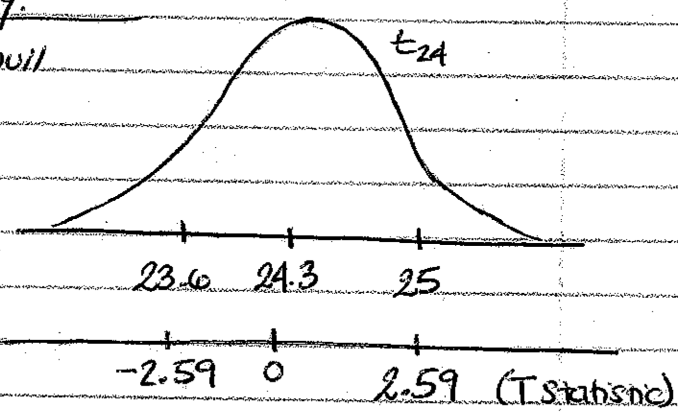
T-Test for Crab body T°C.

$$\frac{\bar{Y} - \mu_0}{(s/\sqrt{n})} = \frac{25^\circ c - 24.3^\circ c}{(0.27)} = \boxed{\frac{\text{signal}}{\text{noise}}} = 2.59 = t$$

2.59 = **T- STATISTIC**

meaning: μ_0 is about 2½ standard errors away from $\bar{Y}$.

Long Run Histogram if null true, accounting for uncertainty in σ



What is the chance of getting data this extreme?

NUMERICAL SURPRISE MEASURE } **P Value** → The Area to the Right of your T statistic

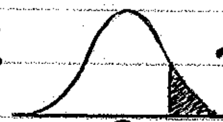
As T↑, P↓

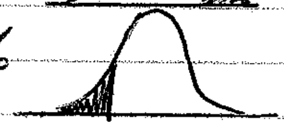
- a larger P value means that the data is not as surprising (high chance of getting such data).

Maya Sertic
2/17/09
2/1

THE DIFFERENT ALTERNATIVES:

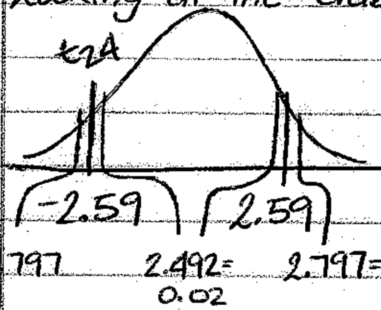
Alternative 1: $\mu \neq \mu_0$ $\longrightarrow$ 2-sided alternative
2-tailed P-value
(2-tailed test)

Alternative 2: $\mu > \mu_0$


Alternative 3: $\mu < \mu_0$


$\longrightarrow$ 1-sided alternative
1-tailed P-value
(1-tailed test)

Looking at the crab data:



On T-Table, look at 24 degrees of freedom, and find where the calculated p. value fits

- Our particular p value of 2.59 is not on the T-table, but it is somewhere between the values of $\underbrace{2.797}_{0.01}$ & $\underbrace{2.492}_{0.02}$

Area in 2 Tails:

Our p-value is between 1% & 2% = small
Small p value $\rightarrow$ reject

We appeal to conventional values to decide if p is small is small enough to reject null & favor the alternative.

$p \leq 5\%$, reject null : result = STATSIG
$p \leq 1\%$, result = highly STATSIG

We estimated our P-value to be between 1% & 2%.
We can use JMP to get a finer value:
2.59 $\rightarrow$ 1.6% (STATSIG, but not highly statsig)

When hypothesis testing is done with a 2-sided alternative, its conclusion is identical to the CI approach.

Maya Sertic

2/17/09

3/7

PITFALLS OF HYPOTHESIS TESTING

1) more complicated, more vocabulary, not as transparent as CI.

2) Usually hypothesis testing is done the opposite way:

null = theory is wrong

alternative = theory is right

because it is easier to manipulate the P value. Most scientific journals want $P \leq 5\%$, so if your 2-tailed P-value is too big, you can just cut it in half and publish with a 1-tailed P-value

1-tailed $p = 4\%$ $\boxed{=}$ 2 tailed $p = 8\%$

Rigid adherence to $p \leq 5\%$ is silly.

3) For a 2-tailed P-value, you cannot tell if $\bar{y}$ is above or below μ_0 (cannot reconstruct $\bar{y}$ from the result, or any thing else for that matter).

4) It is useful to know how big the signal $(\bar{y} - \mu_0)$ and the noise $(s/\sqrt{n})$ are, but if you get:

$$\frac{\text{signal}}{\text{noise}} = 0.5$$

You don't know if 0.5 stands for $\frac{1}{2}$ or $\frac{2}{4}$, and because of this (once again) you can't reconstruct $\bar{y}$

- With CI you can back-calculate to find out what the original #s were

ex: $\bar{y}$ is right in the middle of the interval.

5) For every new theoretical μ_0, you need to build a new hypothesis test (you can reuse same CI).

Maya Sertic 02/17/09

PITFALL COMMON TO HYPOTHESIS TEST & CI

- STATSIG ≠ PRACTSIG
  - Sometimes it turns out that these two are out of focus in relation to each other, so we may conclude that a difference is practsig but not statsig, or the other way around.

EXAMPLE: Systolic Blood Pressure.

There is a new drug out that is meant to lower Systolic Blood Pressure. The results of a test are: on 1000 patients

$\bar{Y} = -1\text{mmHg}$

$S = 10\text{mmHg}$

NOT PRACTSIG, because a blood pressure that is lowered by only 1mmHg does not mean anything to the health of a human — it is much like a blood pressure change of 0mmHg.

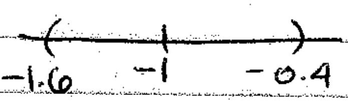
- not Clinically/medically significant

Statsig?

$\mu_0 = 0$ (-1mmHg does nothing, is like 0)

$$95\%\ \text{CI for } \mu = \bar{Y} \pm (t_{999}^{0.95})(S/\sqrt{n})$$
$$= -1 \pm (1.96)(10/\sqrt{1000})$$
$$= -1.6\ \&\ -0.4$$

$t = Z$ because n is too big for T-table ⇒ use normal table.



0mmHg is not in interval ⇒ STATSIG

The data comes out **STATSIG** but **not practsig** because there was **too much data**.

- Standard Error got really small because sample size n was very big.

Maya Sertic

2/17/09

GENERALLY A GOOD WAY TO MEASURE PRACTICAL SIGNIFICANE:

Compute the difference between $\bar{y}$ & μ_0 in relation to μ_0

$\bar{y}$ = treatment

μ_0 = control

$$100\left(\frac{T-C}{C}\right)\% \quad \left[\geq 5\% = \text{practsig}\right]$$

The relative difference on an order of 5% or more tends to be large in practical terms (practsig)

EXAMPLE: Does Calcium concentration in an arthropod species match the Calcium concentration ($[Ca^{2+}]$) in the water?

$[Ca^{2+}]$ of water = 32 mmole/kg

Sample size n = 4 arthropods

$\bar{y}$ $[Ca^{2+}]$ = 29.8 mmole/kg

S = 1.79 mmole/kg

PRACTICALLY SIGNIFICANT?

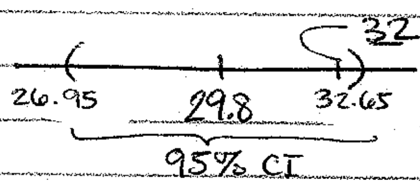
$$100\left(\frac{T-C}{C}\right)\% = 100\left(\frac{29.8-32}{32}\right)\% = 6.9\%$$

PRACTSIG

STATSIG?

build a 95% CI for $\mu = \bar{y} \pm (t_3^{0.95})(S/\sqrt{n})$

$= 29.8 \pm (3.182)(1.79/\sqrt{4}) = 26.95$ & 32.65



32 does fit in the 95% CI

NOT STAT SIG

NOT ENOUGH DATA: Practsig, not Statsig

The SE is too big

TOO MUCH DATA: not Practsig, yes Statsig

The SE is too small

MS

02/17/09

6/7

How many arthropods need to be collected for practical significance & statistical significance to be in focus?

Build a 95% CI w/ n observations

95% CI = 100 $(1-\alpha)$% → $\boxed{\alpha = 0.05}$

$$\bar{y} \pm \left(t_{n-1}^{0.95(2)}\right)\left(S/\sqrt{n}\right)$$

→ area in 2 tails combined must be 5%

This is a pre-experiment calculation

- You need to have some past experience in a similar experiment to have a vague idea of where to start.
  - Suppose past experience makes you think S = 1.8.

  Somebody told you a $[Ca^{2+}]$ of 31.5 in an arthropod is a lot different from a $[Ca^{2+}] = 32$.

SAMPLE SIZE DETERMINATION

CI will discriminate between 2 theories:

$\mu_0 = 32$ mmol/kg

$\mu_A = 31.5$ mmol/kg

If μ_0 falls just outside the 95% CI:

$$\mu_0 = \mu_A + \left(t_{n-1}^{0.95(2)}\right)\left(S/\sqrt{n}\right) \rightarrow \text{SOLVE FOR } \underline{n}$$

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

PROBLEM: T depends on n → n technically on both sides of equation

Equation must be solved iteratively such that your initial guess for n = ∞

$n = \infty$ → follows normal curve

$t = Z = 1.96$ for 95% certainty

* You can trick JMP into doing iterations for you.

MS

02/17/09

7/7

ARTHROPODS → how many?

$s = 1.8$ mmol/kg

$(M_0 - M_A) = (32 - 31.5) = 0.5$ mmol/kg

$$n = \frac{(t_{n-1}^{(1-\alpha)2})^2 (s)^2}{(M_0 - M_A)^2} \qquad n_i = \infty$$

$$\infty = \frac{(1.96)^2(1.8)^2}{(0.5)^2} = 49.8 = 50 = n$$

$n-1 =$ in T table, 49 degrees of freedom ⇒ 2.009

$$50 = \frac{(2.009)^2(1.8)^2}{(0.5)^2} = 52.3 \approx 53 = n$$

52 → 2.008

$$53 = \frac{(2.008)^2(1.8)^2}{(0.5)^2} = 52.3 \approx 53$$

53 is repeated → $\boxed{n = 53}$

Our sample size should be about 53 arthropods to be sure that our practical significance and statistical significance are focused in relation to each other.