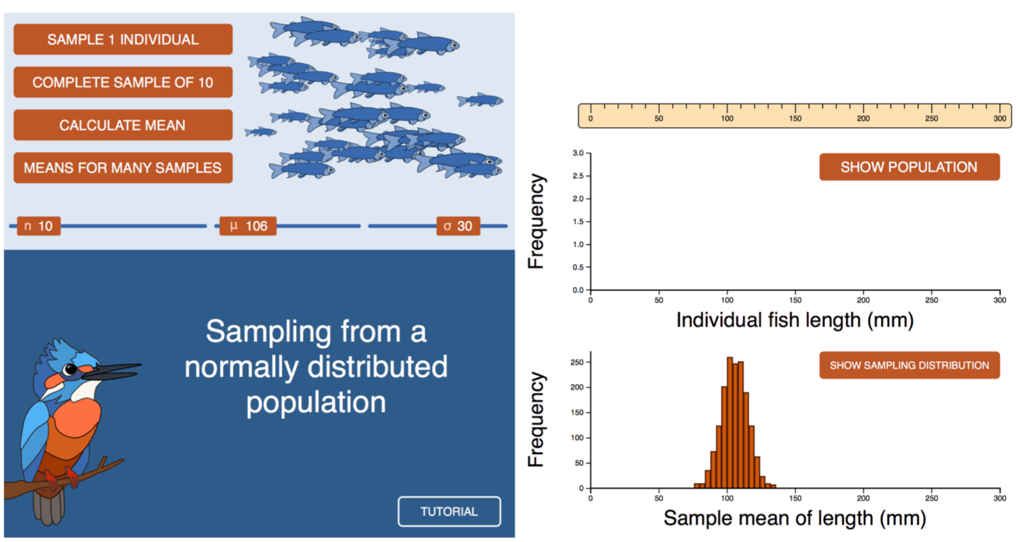
Chapter 11:
Inference for a normal population

Distribution of sample means

- Recall that $\overline{Y}$ has a sampling distribution that is normal



Distribution of sample means

- Recall that $\overline{Y}$ has a sampling distribution that is normal
- Z-standardization for distribution of sample means?

$$Z = \frac{\overline{Y} - \mu}{\sigma_{\overline{Y}}} \quad \text{where} \quad \sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}}$$

But rarely know true population standard deviation

Can estimate/hypothesize μ (true population mean)

Distribution of sample means

- Recall that $\overline{Y}$ has a sampling distribution that is normal
- *Z*-standardization for distribution of sample means?

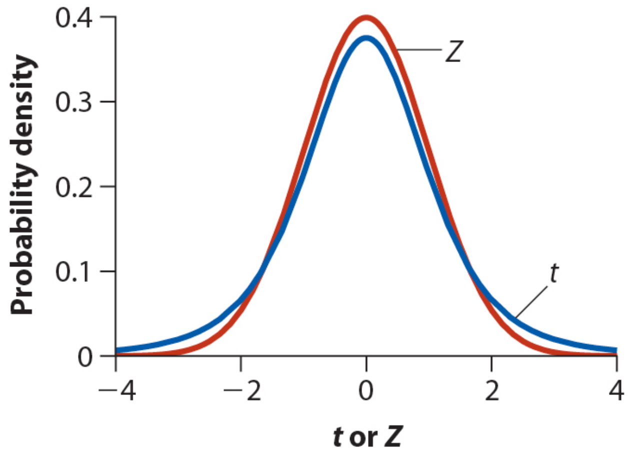
$$Z = \frac{\overline{Y} - \mu}{\sigma_{\overline{Y}}} \quad \text{where} \quad \sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}}$$

substitution

$$t = \frac{\overline{Y} - \mu}{SE_{\overline{Y}}} \quad \text{where} \quad SE_{\overline{Y}} = \frac{s}{\sqrt{n}}$$

Student's t distribution with *n* – 1 degrees of freedom

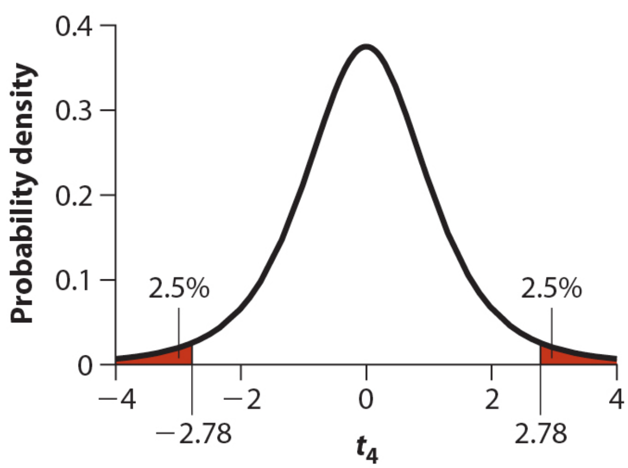
Student's t distribution



- Similar to standard normal distribution (Z), but with fatter tails
- The t-distribution shown here has *df* = 4
- As the sample size increases the t distribution becomes more like the standard normal distribution

Fig 11.1-1

Student's t distribution



Has critical values like Z

Get values with computer or table

Fig 11.1-2

Student's t distribution

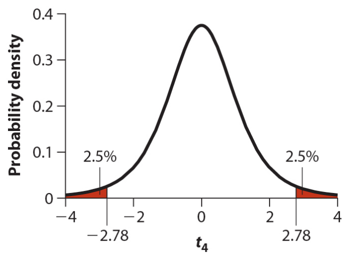


TABLE 11.1-1 Critical values of the *t*-distribution. Excerpted from Statistical Table C.

df	$\alpha(2) = 0.20$ $\alpha(1) = 0.10$	$\alpha(2) = 0.10$ $\alpha(1) = 0.05$	$\alpha(2) = 0.05$ $\alpha(1) = 0.025$	$\alpha(2) = 0.02$ $\alpha(1) = 0.01$	$\alpha(2) = 0.01$ $\alpha(1) = 0.005$
1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	**2.78**	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03
...	...	...	...	...	...

Fig 11.1-2

Ex 11.2: Eye to eye

- Can use t-distribution to accurately calculate a confidence interval for the mean of a population with a normal distribution
- Stalk-eyed flies
- Sample: 8.69, 8.15, 9.25, 9.45, 8.96, 8.65, 8.43, 8.79, 8.63
- $\overline{Y}$ = 8.778 and s = 0.398

Ex 11.2: Eye to eye

- Sample: 8.69, 8.15, 9.25, 9.45, 8.96, 8.65, 8.43, 8.79, 8.63
- $\overline{Y}$ = 8.778 and s = 0.398
- 95% confidence interval for the population mean

$$SE_{\overline{Y}} = \frac{s}{\sqrt{n}}$$

$$\overline{Y} - \underbrace{t_{0.05(2),df}}_{\text{critical value from table/computer}} SE_{\overline{Y}} < \mu < \overline{Y} + t_{0.05(2),df} SE_{\overline{Y}}$$

$$SE_{\overline{Y}} = \frac{s}{\sqrt{n}} = \frac{0.398}{\sqrt{9}} = 0.133$$

$$t_{0.05(2),8} = 2.31$$

$$8.778 - (2.31 \times 0.133) < \mu < 8.778 + (2.31 \times 0.133)$$

$$8.47 < \mu < 9.08$$

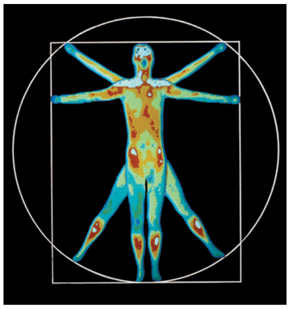
One-sample t-test

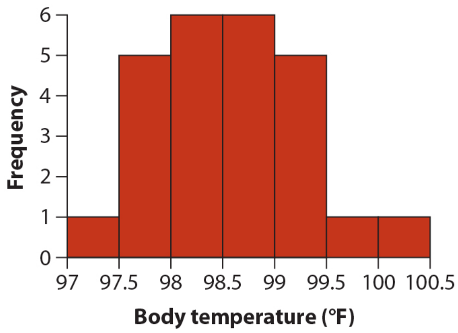
- The **one sample t-test** compares the mean of a random sample from a normal population with the population mean proposed in a null hypothesis
- H_0: the true mean equals μ_0
- H_A: the true mean does not equal μ_0
- Test statistic:

$$t = \frac{\overline{Y} - \mu_0}{SE_{\overline{Y}}} \qquad \textit{where} \qquad SE_{\overline{Y}} = \frac{s}{\sqrt{n}}$$

$$df = n - 1$$

Ex 11.3: Human body temperature





- Normal = 98.6ºF
- Is this really normal?
- Sample of 25 healthy people
- One-sample t-test
  - H_0: mean body temp equals 98.6ºF
  - H_A: mean body temp does not equal 98.6ºF

Fig 11.3-1

Ex 11.3: Human body temperature

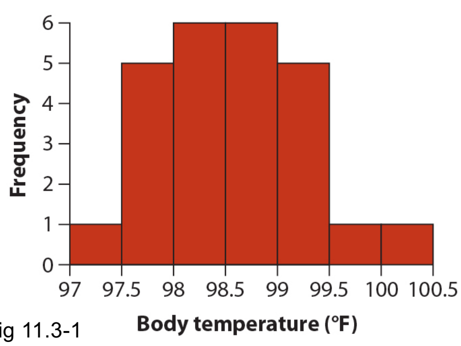


Fig 11.3-1

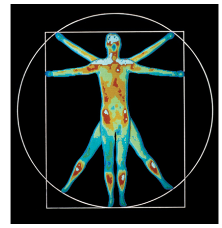
- From the sample of 25
  - $\overline{Y}$ = 98.524
  - s = 0.678

$$SE_{\overline{Y}} = \frac{s}{\sqrt{n}} = \frac{0.678}{\sqrt{25}} = 0.136$$

$$t = \frac{\overline{Y} - \mu_0}{SE_{\overline{Y}}} = \frac{98.524 - 98.6}{0.136} = -0.56$$

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

Interpreting t-statistic



$t_{24} = -0.56$

- Compute *P*-value: probability of this t-statistic (or more extreme) given the null hypothesis (μ = 98.6) is true

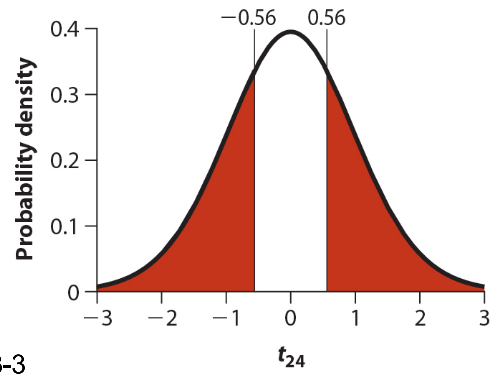


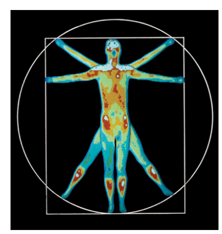
Fig 11.3-3

P = Pr[t < -0.56] + Pr[t > 0.56]

Using a computer:

- *P* = 0.58
- Fail to reject the null hypothesis

Interpreting t-statistic



$t_{24} = -0.56$

- Compute *P*-value: probability of this t-statistic (or more extreme) given the null hypothesis (μ = 98.6) is true

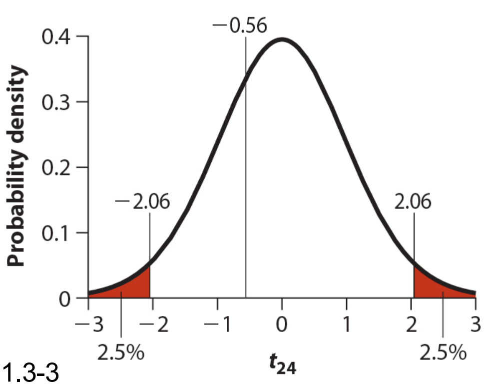
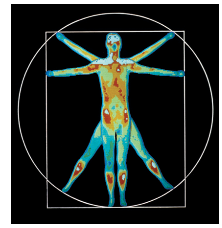


Fig 11.3-3

P = Pr[t < -0.56] + Pr[t > 0.56]

Using a stats table

- Look up critical t-value $t_{0.05(2),24}$
- Critical t-value = 2.06
- Observed value is within range of -/+ critical value
- Data consistent with true null



Is common wisdom correct?

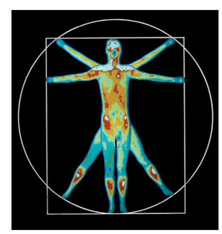
- How much uncertainty is there in our estimate of μ?
- 95% confidence interval

$$\overline{Y} - t_{0.05(2),df} SE_{\overline{Y}} < \mu < \overline{Y} + t_{0.05(2),df} SE_{\overline{Y}}$$

$$98.524 - (2.06 \times 0.136) < \mu < 98.524 + (2.06 \times 0.136)$$

$$98.24 < \mu < 98.80$$

Value of 98.6 is within 95% CI (as seen by P),
but slightly different values are also consistent with these data



Increasing sample size

- Increasing sample size reduces standard error of mean
  - Uncertainty of estimate of mean
- Increase sample size from 25 to 130
- $t_{129} = -5.44; P = 0.000016$
- 95% CI = $98.12 < \mu < 98.38$
- Larger sample sizes increase probability of rejecting a false null hypothesis (power)
- If this null is really false, then the sample of 25 failed to detect a false null (Type II error)

Assumptions of one-sample t-test

- Data are a random sample from the population
- Variable is normally distributed in the population
  - Few variables in biology are exact match to normality
  - But in many cases the test is robust to departures from normality (more on this later)

Estimating other statistics

- Emphasis on estimating the mean of a normal population
- How about other statistics?
  - Spread of the sample distribution (standard deviation or variance)
- Confidence limits for variance is based on the χ^2 distribution

Confidence interval for variance

$$\frac{df\ s^2}{\chi^2_{\alpha/2,df}} < \sigma^2 < \frac{df\ s^2}{\chi^2_{1-\alpha/2,df}}$$

Confidence interval for standard deviation

$$\sqrt{\frac{df\ s^2}{\chi^2_{\alpha/2,df}}} < \sigma < \sqrt{\frac{df\ s^2}{\chi^2_{1-\alpha/2,df}}}$$

Assumptions of calc confidence intervals for variance

- Random sample from the population
- Variable must have normal distribution
  - Formulas are NOT robust to departures from normality