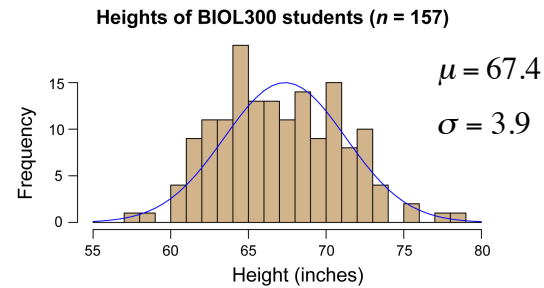
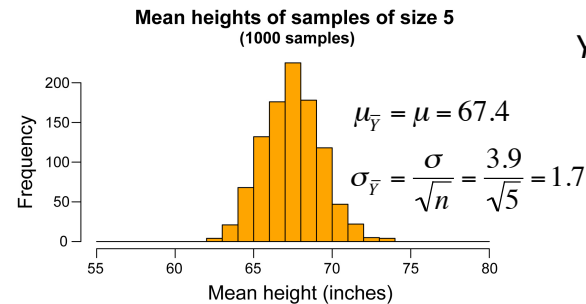


Numerical variables from a single sample

Chapter 11



\bar{Y} is normally distributed
whenever:
 Y is normally distributed
or
 n is large



Inference about means

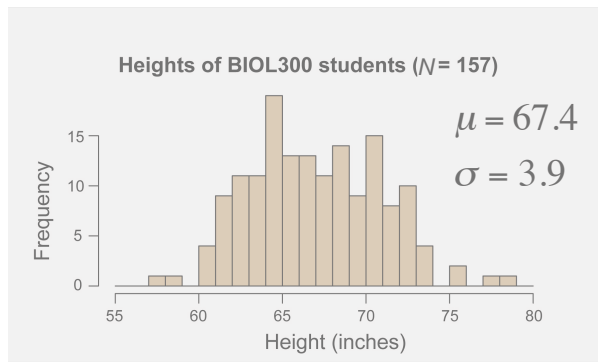
Because \bar{Y} is normally distributed, we can convert its distribution to a standard normal distribution:

$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}} = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}$$

This would give a probability distribution of the difference between a sample mean and the population mean.

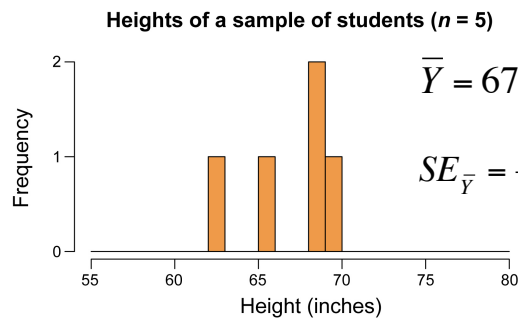
But... We don't know σ ...

However, we do know s , the standard deviation of our sample. We can use that as an estimate of σ .



In most cases, we don't know the real population distribution.

We only have a sample.



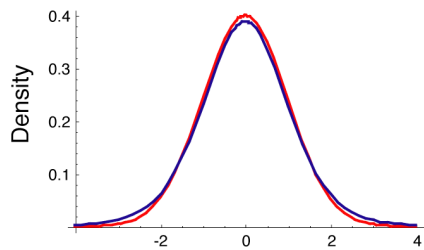
We use this as an estimate of $\sigma_{\bar{Y}}$

A good approximation to the standard normal is then:

$$t = \frac{\bar{Y} - \mu}{SE_{\bar{Y}}} = \frac{\bar{Y} - \mu}{s/\sqrt{n}}$$

t has a Student's t distribution

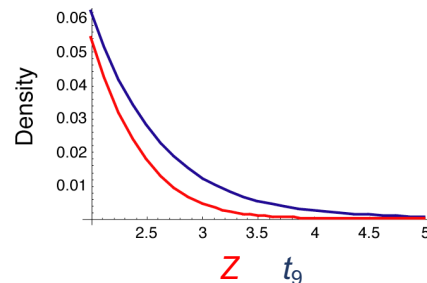
Degrees of freedom



Discovered by William Gossett, of the Guinness Brewing Company

$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}}$$

$$t = \frac{\bar{Y} - \mu}{SE_{\bar{Y}}}$$



$$df = n - 1$$

We use the t -distribution to calculate a confidence interval of the mean

$$-t_{\alpha(2),df} < \frac{\bar{Y} - \mu}{SE_{\bar{Y}}} < t_{\alpha(2),df}$$

We rearrange the above to generate:

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

Another way to express this is: $\bar{Y} \pm SE_{\bar{Y}} t_{\alpha(2),df}$

<https://www.youtube.com/watch?v=16aGSx9gFO4>

95% confidence interval for a mean

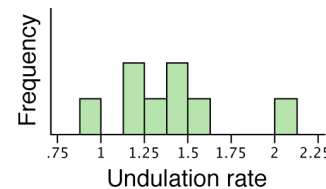
Example:
Paradise flying snakes



Undulation rates (in Hz)

0.9, 1.4, 1.2, 1.2, 1.3, 2.0, 1.4, 1.6

Estimate the mean and standard deviation



$$\bar{Y} = 1.375$$

$$s = 0.324$$

$$n = 8$$

Find the standard error

$$\bar{Y} \pm SE_{\bar{Y}} t_{\alpha(2), df}$$

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}} = \frac{0.324}{\sqrt{8}} = 0.115$$

Find the critical value of t

$$df = n - 1 = 7$$

$$t_{\alpha(2), df} = t_{0.05(2), 7} = 2.36$$

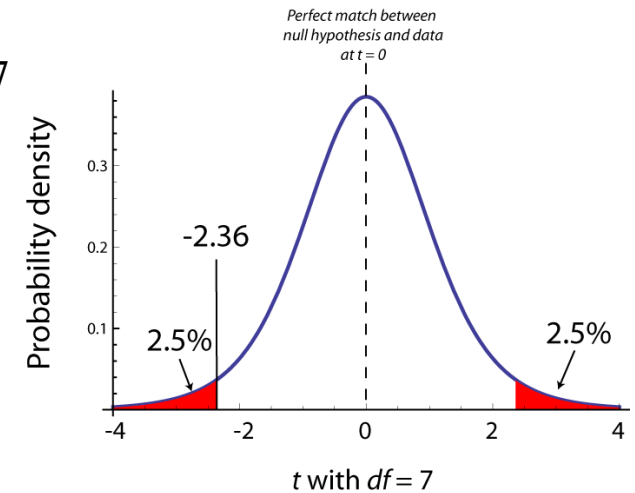


Table C: Student's t distribution

	$\alpha(2)$:	0.2	0.10	0.05	0.02	0.01	0.001	0.0001
df	$\alpha(1)$:	0.1	0.05	0.025	0.01	0.005	0.0005	0.00005
1		3.08	6.31	12.71	31.82	63.66	636.62	6366.20
2		1.89	2.92	4.30	6.96	9.92	31.60	99.99
3		1.64	2.35	3.18	4.54	5.84	12.92	28.00
4		1.53	2.13	2.78	3.75	4.60	8.61	15.54
5		1.48	2.02	2.57	3.36	4.03	6.87	11.18
6		1.44	1.94	2.45	3.14	3.71	5.96	9.08
7		1.41	1.89	2.36	3.00	3.50	5.41	7.88
8		1.40	1.86	2.31	2.90	3.36	5.04	7.12
9		1.38	1.83	2.26	2.82	3.25	4.78	6.59

Finding the critical value of t in R

```
qt(0.025, df = 7, lower.tail = FALSE)
[1] 2.364624
```

Putting it all together...

$$\begin{aligned}\bar{Y} \pm SE_{\bar{Y}} t_{\alpha(2),df} &= 1.375 \pm 0.115 \quad (2.36) \\ &= 1.375 \pm 0.271\end{aligned}$$

$$1.10 < \mu < 1.65$$

(95% confidence interval)

Confidence interval for mean
in R

```
t.test(undulationRate)$conf.int  
[1] 1.104098 1.645902  
attr(,"conf.level")  
[1] 0.95
```

```
t.test(undulationRate,  
       conf.level = 0.99)$conf.int  
[1] 0.9740838 1.7759162  
attr(,"conf.level")  
[1] 0.99
```

99% confidence interval

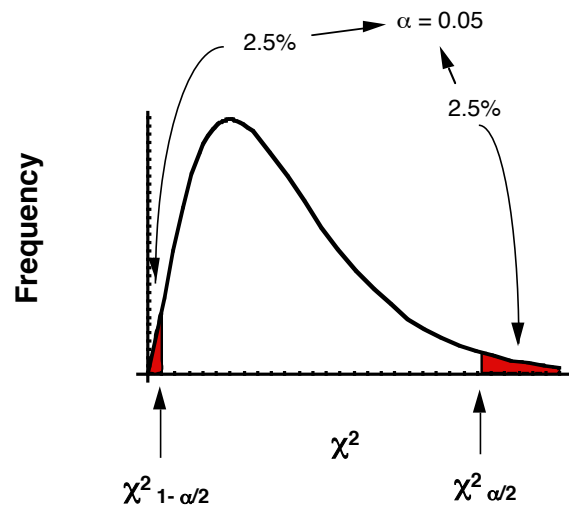
$$t_{\alpha(2),df} = t_{0.01(2),7} = 3.50$$

$$\begin{aligned}\bar{Y} \pm SE_{\bar{Y}} t_{\alpha(2),df} &= 1.375 \pm 0.115 \quad (3.50) \\ &= 1.375 \pm 0.403\end{aligned}$$

$$0.97 < \mu < 1.78$$

Confidence interval for the variance

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



95% confidence interval for the variance of flying snake undulation rate

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

$$df = n - 1 = 7$$

$$s^2 = (0.324)^2 = 0.105$$

$$\chi^2_{\frac{\alpha}{2}, df} = \chi^2_{0.025, 7} = 16.01$$

$$\chi^2_{1-\frac{\alpha}{2}, df} = \chi^2_{0.975, 7} = 1.69$$

Table A

df	0.999	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005	0.001
X										
1	1.6 E-6	3.9E-5	0.00016	0.00098	0.00393	3.84	5.02	6.63	7.88	10.83
2	0	0.01	0.02	0.05	0.1	5.99	7.38	9.21	10.6	13.82
3	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84	16.27
4	0.09	0.21	0.3	0.48	0.71	9.49	11.14	13.28	14.86	18.47
5	0.21	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75	20.52
6	0.38	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55	22.46
7	0.6	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28	24.32
8	0.86	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95	26.12

```
> qchisq(0.025, df=7, lower.tail=FALSE)
```

```
[1] 16.01276
```

```
> qchisq(0.975, df=7, lower.tail=FALSE)
```

```
[1] 1.689869
```

95% confidence interval for the variance of flying snake undulation rate

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

$$\frac{7 (0.324)^2}{16.01} \leq \sigma^2 \leq \frac{7 (0.324)^2}{1.69}$$

$$0.0459 \leq \sigma^2 \leq 0.435$$

Test statistic for one-sample t -test

$$t = \frac{\bar{Y} - \mu_0}{s / \sqrt{n}}$$

μ_0 is the mean value proposed by H_0

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.

Hypotheses for one-sample t -tests

H_0 : The mean of the population is μ_0 .

H_A : The mean of the population is not μ_0 .

Example: Human body temperature



H_0 : Mean healthy human body temperature is 98.6°F.

H_A : Mean healthy human body temperature is not 98.6°F.

Human body temperature

$$n = 24$$

$$\bar{Y} = 98.28$$

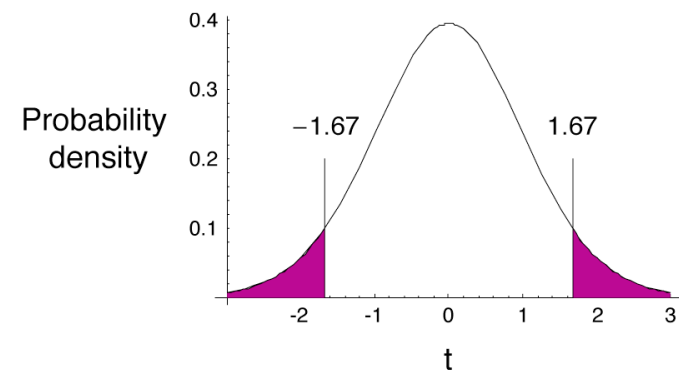
$$s = 0.940$$

$$t = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}} = \frac{98.28 - 98.6}{0.940/\sqrt{24}} = -1.67$$

Degrees of freedom

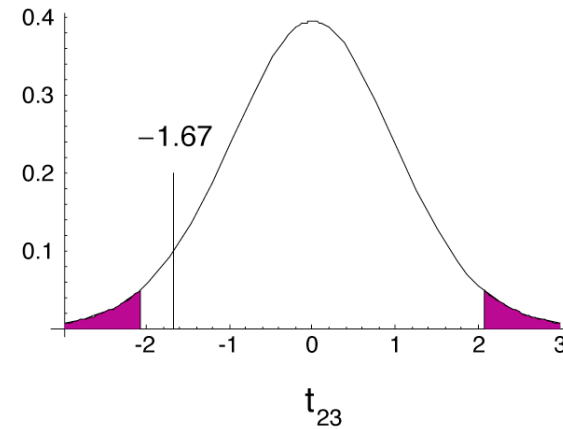
$$df = n - 1 = 23$$

Comparing t to its distribution to find the P -value



A portion of the t table

df	$\alpha(1)$ =0.1 $\alpha(2)=0.2$	$\alpha(1)$ =0.05 $\alpha(2)=0.10$	$\alpha(1)$ =0.025 $\alpha(2)=0.05$	$\alpha(1)$ =0.01 $\alpha(2)=0.02$	$\alpha(1)$ =0.005 $\alpha(2)=0.01$
...
20	1.33	1.72	2.09	2.53	2.85
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.5	2.81
24	1.32	1.71	2.06	2.49	2.8
25	1.32	1.71	2.06	2.49	2.79



-1.67 is closer to 0 than -2.07 , so $P > 0.05$.

With these data, we cannot reject the null hypothesis that the mean human body temperature is 98.6.

```
t.test(bodyTempSmallData$temperature, mu = 98.6)
```

One Sample t-test

```
data: bodyTempSmallData$temperature
t = -0.56065, df = 24, p-value = 0.5802
alternative hypothesis: true mean is not equal to 98.6
95 percent confidence interval:
 98.24422 98.80378
sample estimates:
mean of x
 98.524
```

Body temperature revisited: $n = 130$

$$n = 130$$

$$\bar{Y} = 98.25$$

$$s = 0.733$$

$$t = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}} = \frac{98.25 - 98.6}{0.733/\sqrt{130}} = -5.44$$

Body temperature revisited:
 $n = 130$

$$t = -5.44$$

$$t_{0.05(2),129} = \pm 1.98$$

t is further out in the tail than the critical value, so we could reject the null hypothesis. Human body temperature is not 98.6°F. $P = 2.4 \times 10^{-7}$

One-sample t -test: Assumptions

1. The variable is normally distributed.
2. The sample is a random sample.