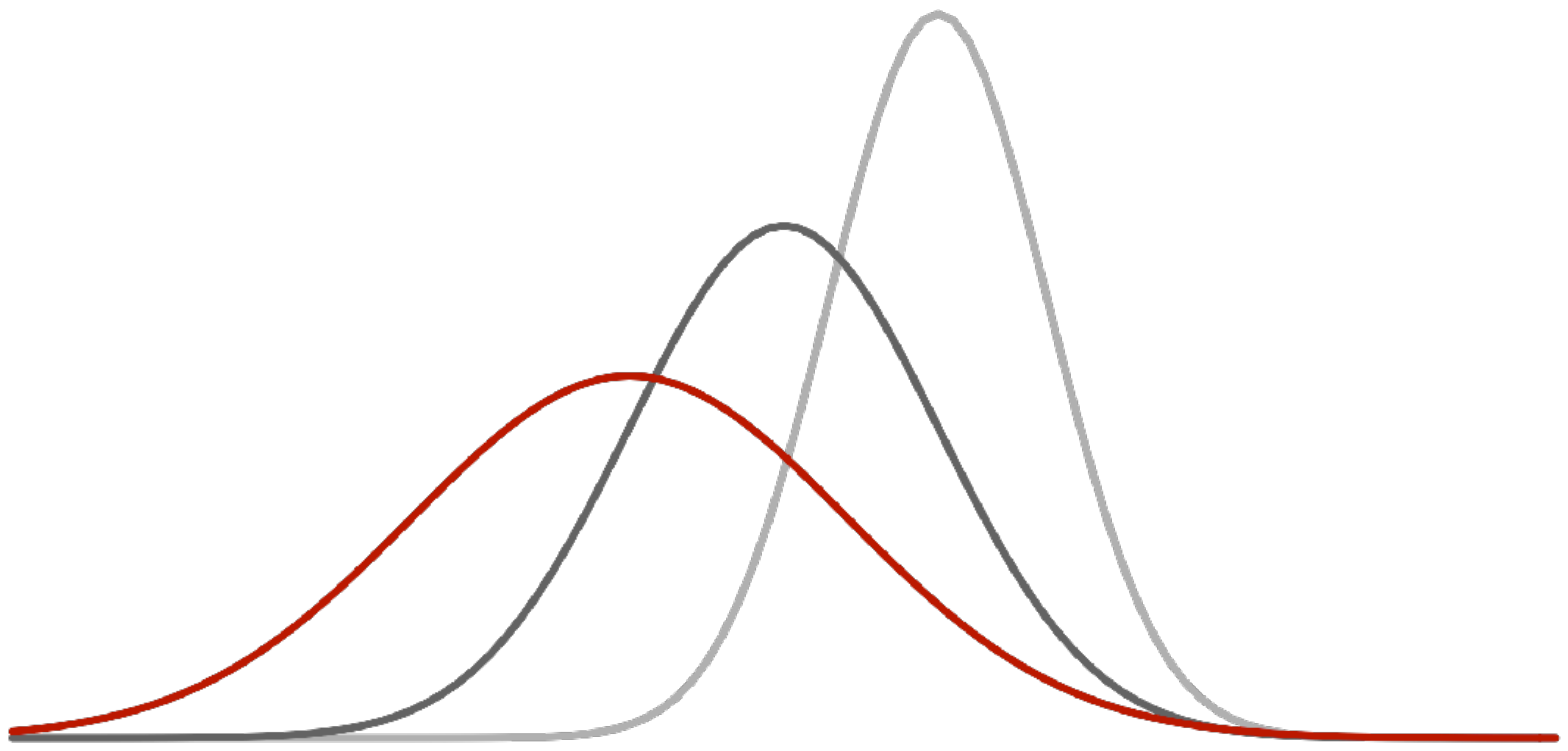
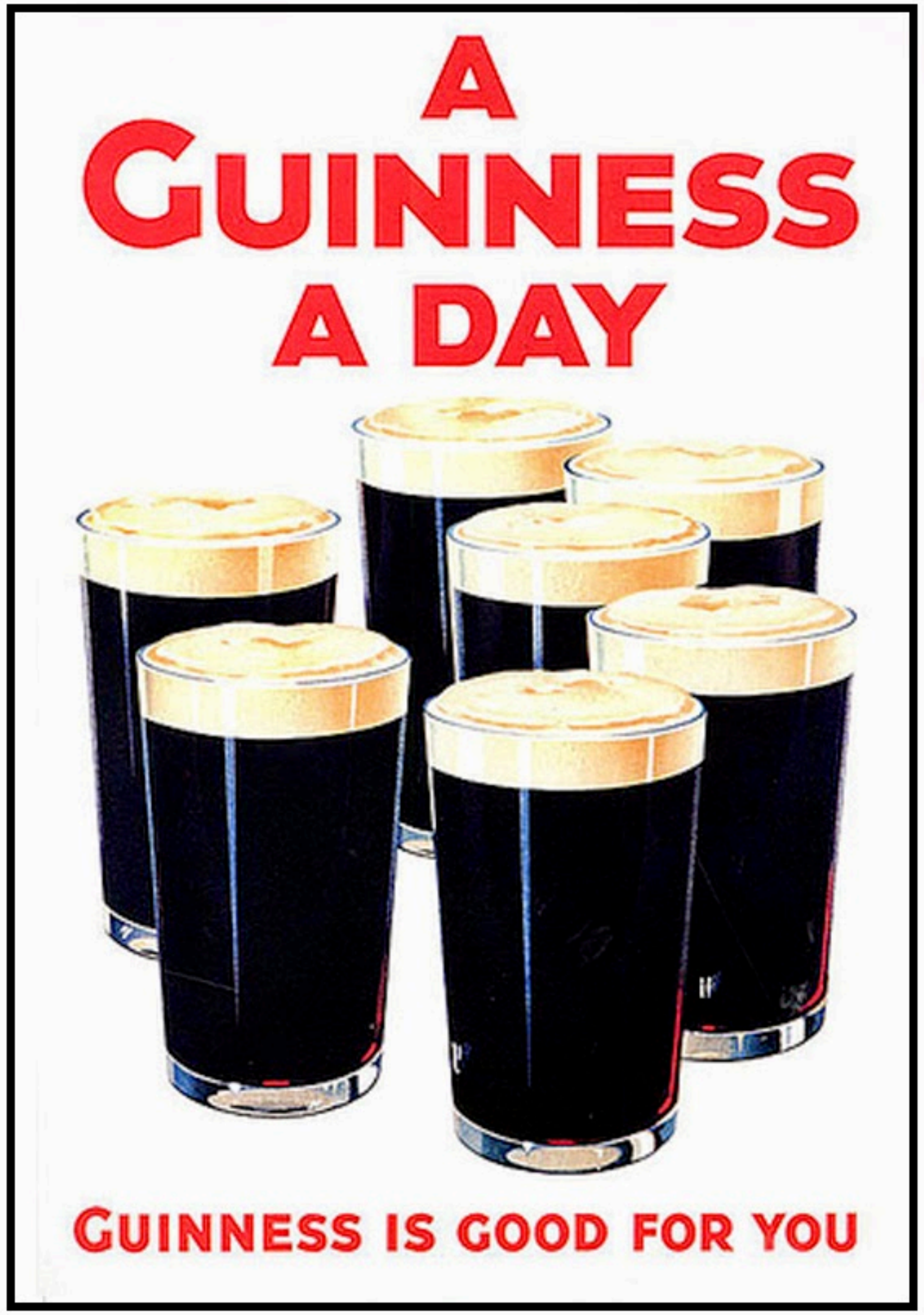


Confidence Intervals

STAT 3202 | OSU | Autumn 2018

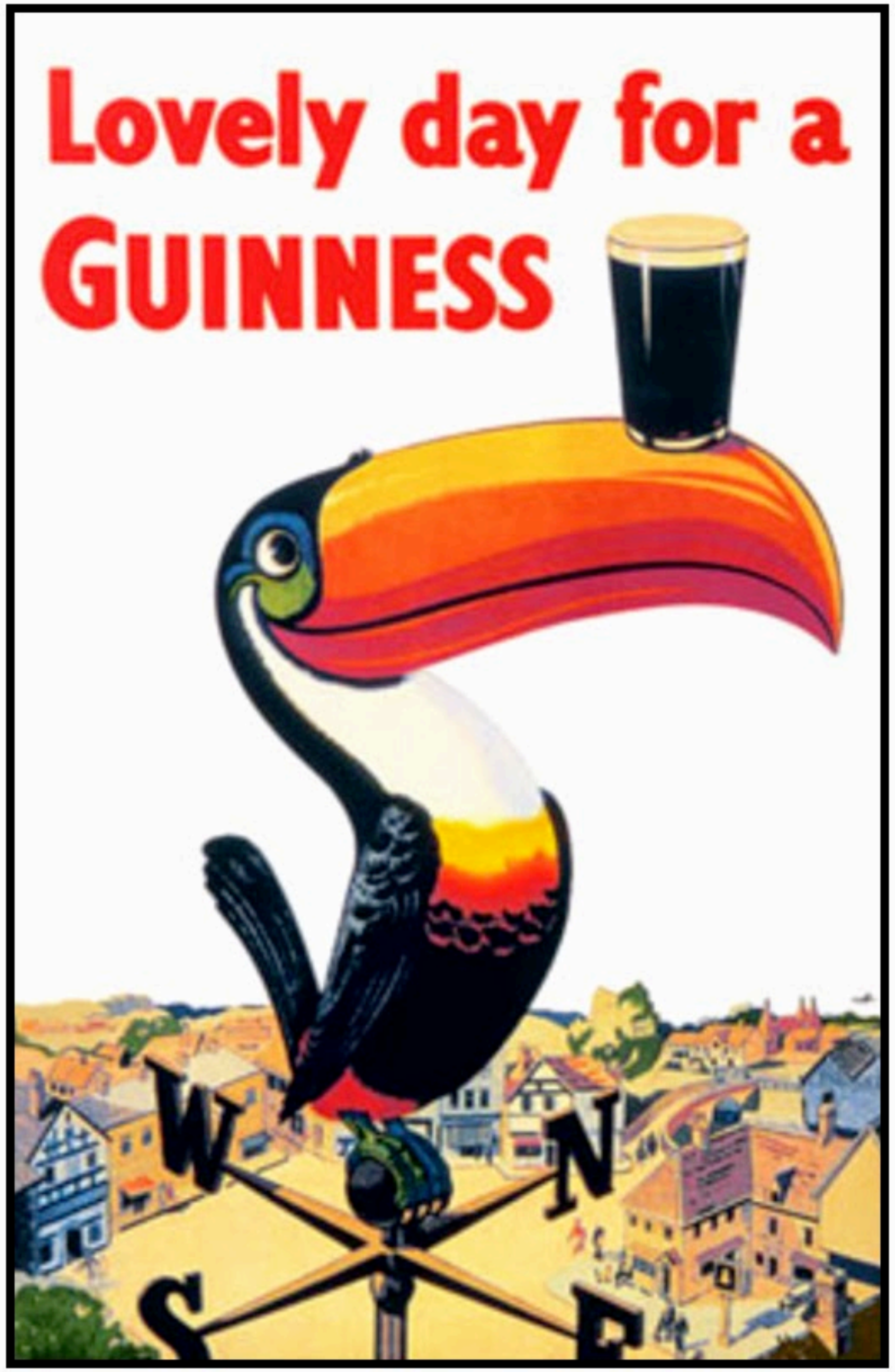
Dalpiaz

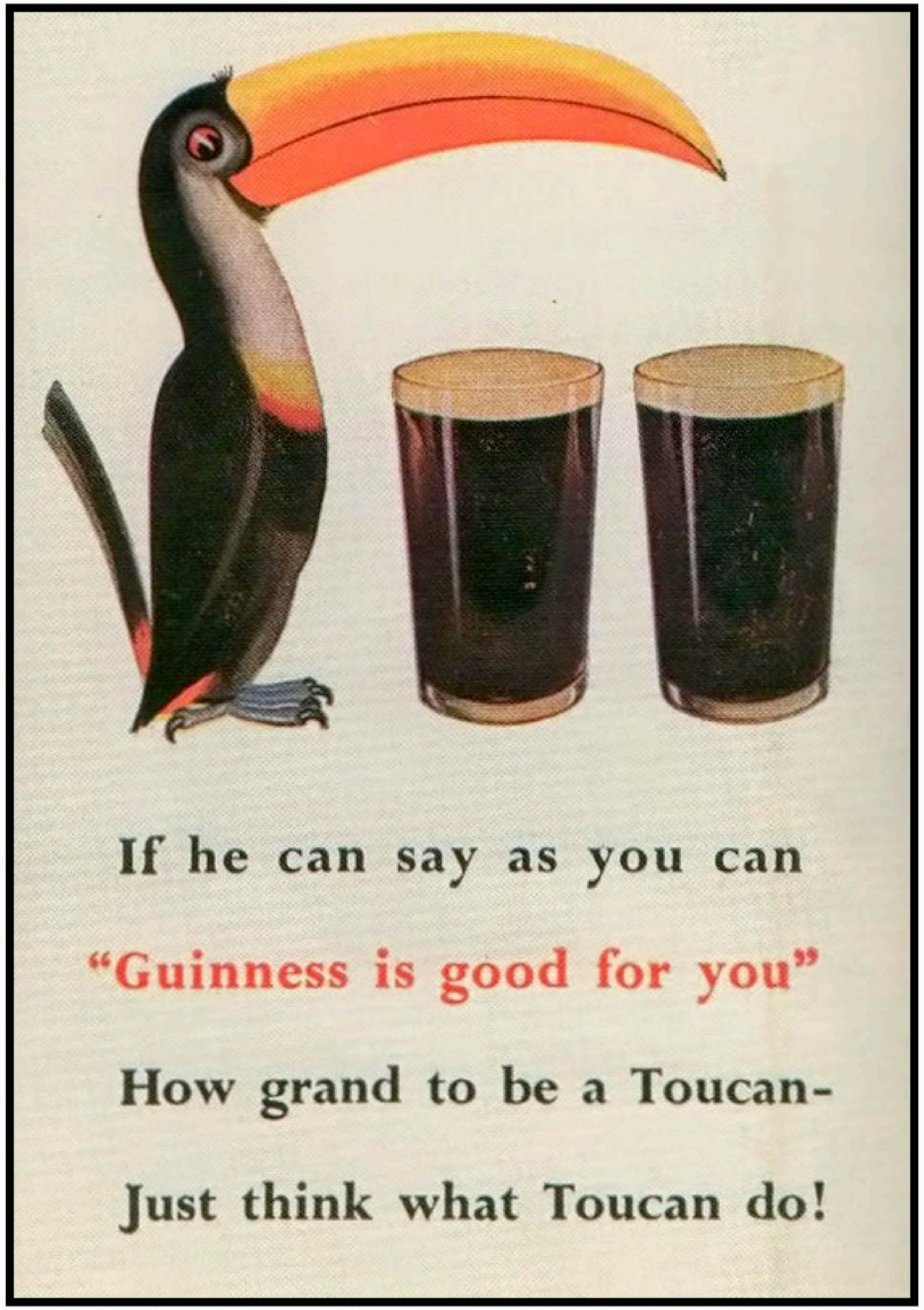

GUINNESS
is good
for you


A
GUINNESS
A DAY
GUINNESS IS GOOD FOR YOU


GUINNESS
FOR STRENGTH

GUINNESS
FOR STRENGTH


Lovely day for a
GUINNESS
W
N
S
E


If he can say as you can
"Guinness is good for you"
How grand to be a Toucan-
Just think what Toucan do!

Confidence Intervals for Means

Let's consider the "usual" setup where we have an (iid) random sample from a normal distribution with parameters μ and σ.

How do we estimate μ?

For a moment, let’s pretend we live in some fantasy land where σ is **known**.

Suppose the lifetime of a particular brand of light bulbs is normally distributed with standard deviation of $\sigma = 75$ hours and unknown mean.

What is the probability that in a random sample of $n = 49$ bulbs, the average lifetime $\overline{X}$ is within 21 hours of the overall average lifetime?

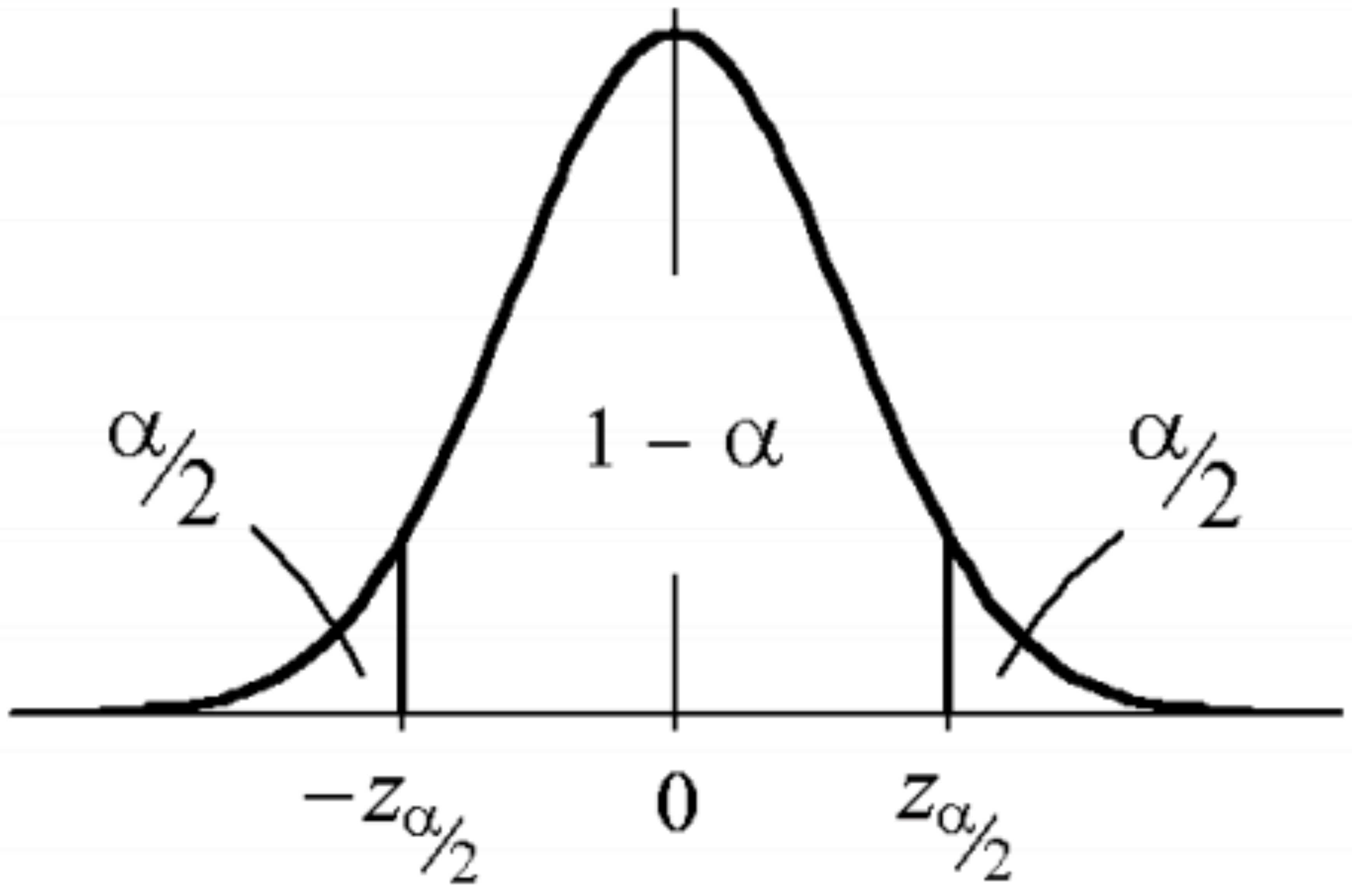
Suppose the lifetime of a particular brand of light bulbs is normally distributed with standard deviation of $\sigma = 75$ hours and unknown mean.

Suppose the sample average lifetime of $n = 49$ bulbs is $\bar{x} = 843$ hours. Construct a 95% confidence interval for the overall average lifetime for light bulbs of this brand.

A **confidence interval** is a *range of numbers* believed to include an unknown population parameter. Associated with the interval is a measure of the *confidence* we have that the interval does indeed contain the parameter of interest.

A $(1 - \alpha)\ 100\%$ confidence interval for the population mean μ when σ is known and sampling is done from a normal population, or with a large sample, is

$$\left(\overline{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\ ,\ \overline{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$



Suppose the sample average lifetime of $n = 49$ bulbs is $\bar{x} = 843$ hours.

Construct a 95% confidence interval for the overall average lifetime for light bulbs of this brand.

Construct a 90% confidence interval for the overall average lifetime for light bulbs.

Construct a 92% confidence interval for the overall average lifetime for light bulbs.

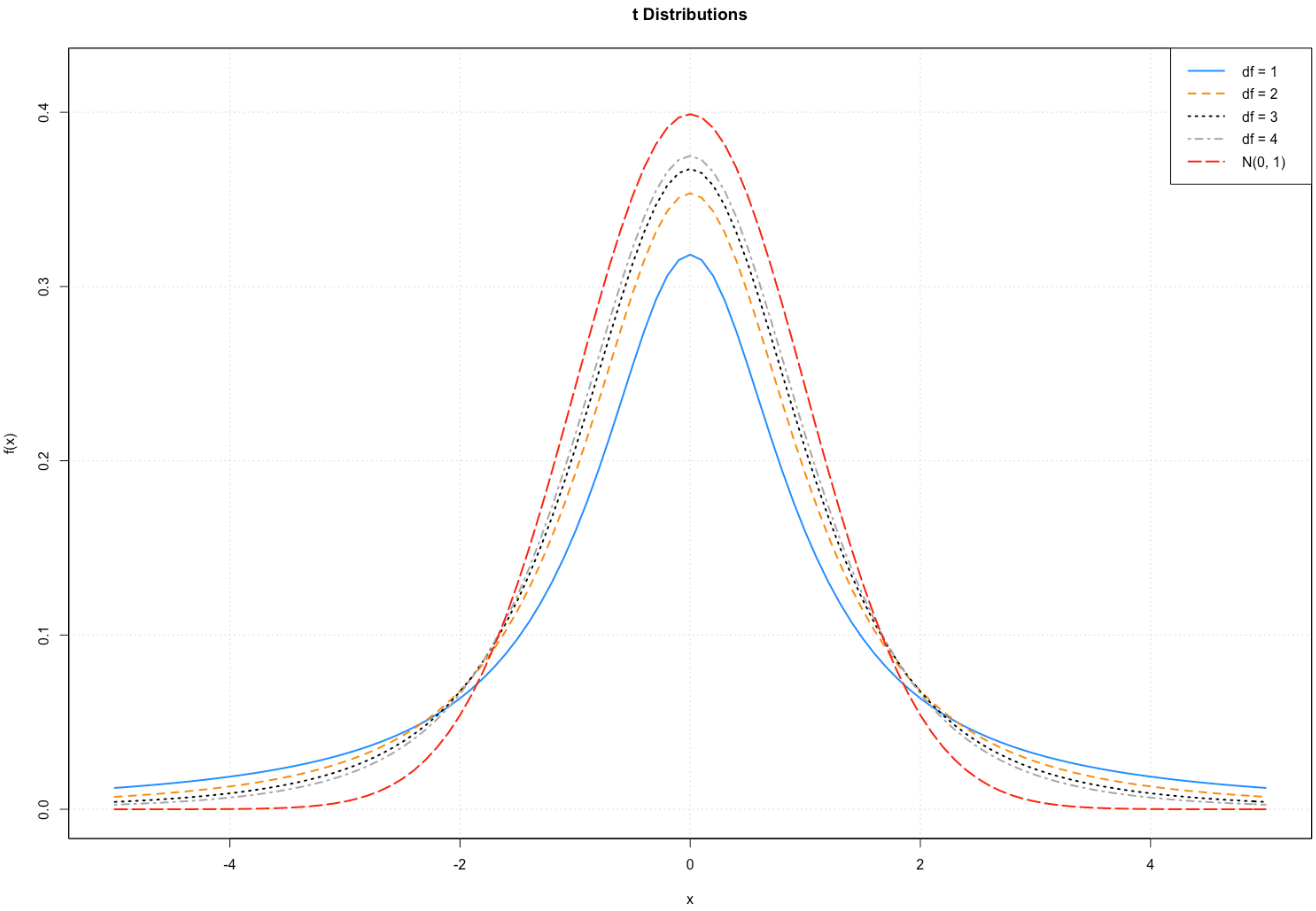
What if we live in reality where σ is **unknown**?

What do we do now?

A $(1-\alpha)\,100\%$ confidence interval for the population mean μ

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \qquad\qquad \bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$n-1$ degrees of freedom



t Distributions
df = 1
df = 2
df = 3
df = 4
N(0, 1)
0.0
0.1
0.2
0.3
0.4
f(x)
-4
-2
0
2
4
x


PEARSON
FISHER
GOSSET
BAYES

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

A manufacturer of TV sets wants to find the average selling price of a particular model. A random sample of 25 different stores gives the mean selling price as \$342 with a sample standard deviation of \$14. Assume the prices are normally distributed. Construct a 95% confidence interval for the mean selling price of the TV model.

The following random sample was obtained from $N(\mu, \sigma^2)$ distribution:

16 12 18 13 21 15 8 17

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

→ Construct a 95% confidence interval for μ.

→ Construct a 90% confidence upper bound for μ.

→ Construct a 99% confidence lower bound for μ.

Confidence Intervals for Proportions

The **sample proportion**:

$$\hat{p} = \frac{x}{n}$$

where x is the number of elements in the sample found to belong to the category of interest (the number of "successes"), and n is the sample size.

$$\mathrm{E}(\hat{P}) = p, \qquad \mathrm{Var}(\hat{P}) = \frac{p \cdot (1-p)}{n}, \qquad \mathrm{SD}(\hat{P}) = \sqrt{\frac{p \cdot (1-p)}{n}}.$$

A large-sample confidence interval for the population proportion p is

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \left(1-\hat{p}\right)}{n}}.$$

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$$

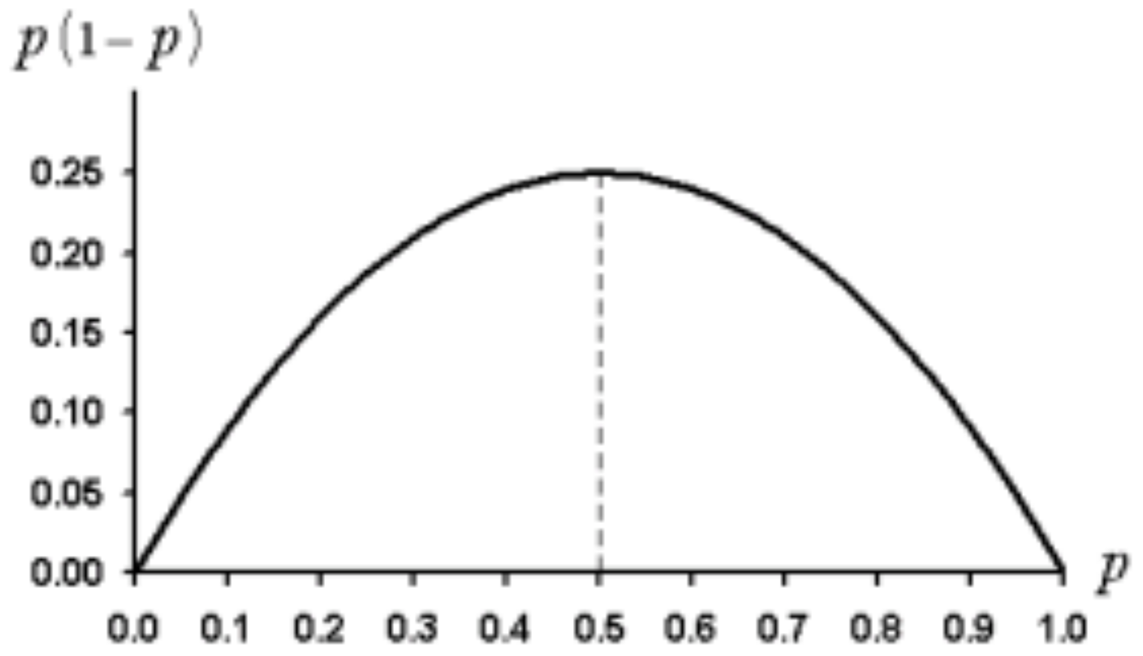
Just prior to an important election, in a random sample of 749 voters, 397 preferred Candidate Y over Candidate Z. Construct a 90% confidence interval for the overall proportion of voters who prefer Candidate Y over Candidate Z.

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \left(1 - \hat{p}\right)}{n}}$$

An article on secretaries' salaries in the *Wall Street Journal* reports: "Three-fourth of surveyed secretaries said they make less than $25,000 a year." Suppose that the *Journal* based its results on a random sample of 460 secretaries drawn from every category of business. Give a 95% confidence interval for the proportion of secretaries earning less than $25,000 a year.

The sample size required to obtain a confidence interval for the population proportion p with specified margin of error ε is

$$n = \left(\frac{z_{\alpha/2}}{\varepsilon} \right)^2 p^* \left(1 - p^* \right).$$



Always round n up.

Conservative Approach: $p^* = 0.50$.

- If it is possible that $p = 0.50$, use $p^* = 0.50$.
- If it is not possible that $p = 0.50$, use p^* = the closest to 0.50 possible value of p.

$$n = \left(\frac{z_{\alpha/2}}{\varepsilon} \right)^2 p^* \left(1 - p^* \right)$$

Find the minimum sample size required for the overall proportion of voters who prefer Candidate Y over Candidate Z to within 2% with 90% confidence. (Assume that no guess as to what that proportion might be is available.)

$$n = \left(\frac{z_{\alpha/2}}{\varepsilon} \right)^2 p^*(1 - p^*)$$

A television station wants to estimate the proportion of the viewing audience in its area that watch its evening news. Find the minimum sample size required to estimate that proportion to within 3% with 95% confidence if …

→ no guess as to the value of that proportion is available.

→ it is known that the station's evening news reaches at most 30% of the viewing audience.

Confidence Intervals for Variances

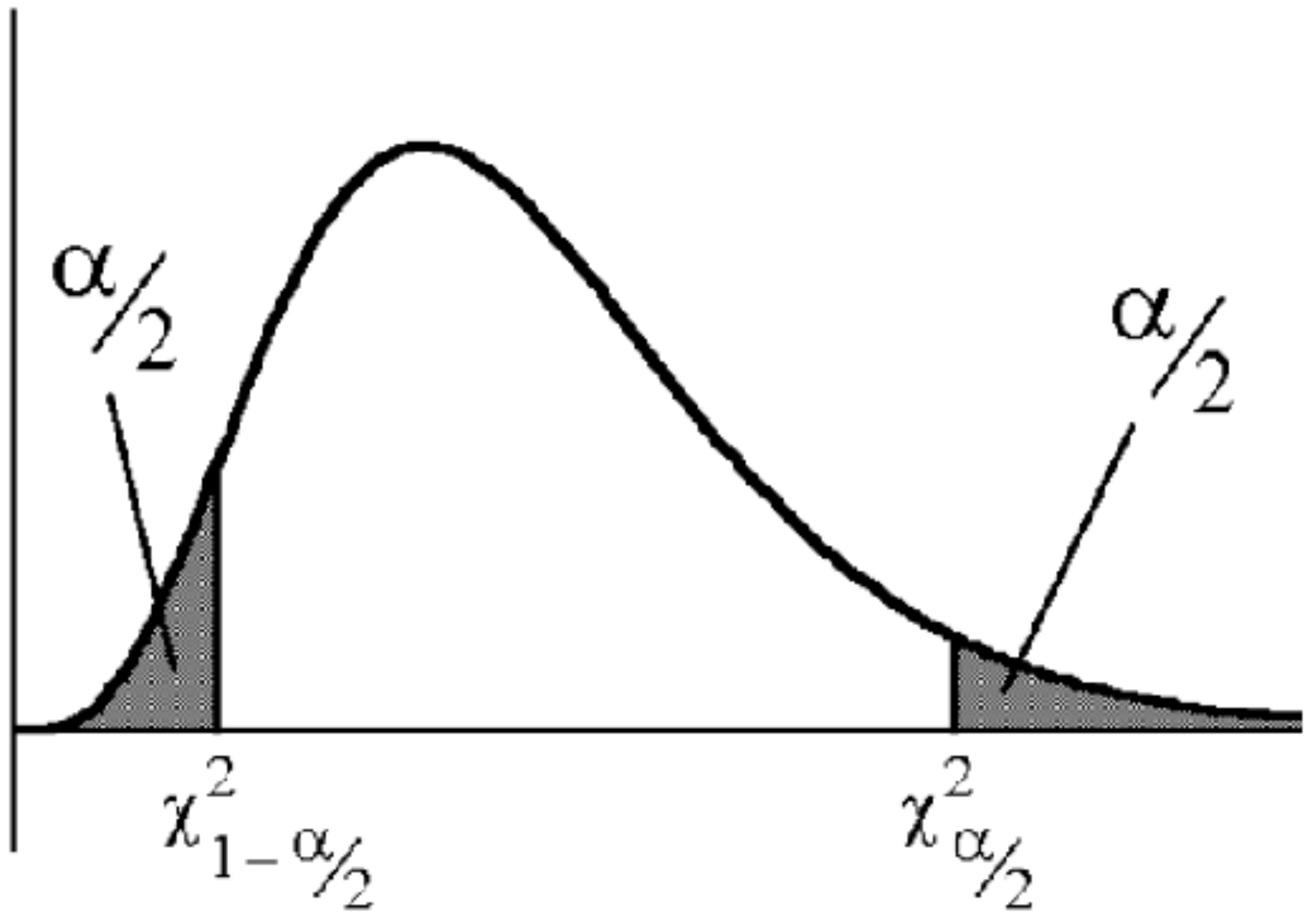
If $X_1, X_2, \ldots, X_n$ are i.i.d. $\mathbf{N}\left(\mu, \sigma^2\right)$. Then

$$\frac{(n-1)\cdot S^2}{\sigma^2} = \frac{\sum\left(X_i - \overline{X}\right)^2}{\sigma^2} \text{ is } \chi^2(n-1).$$

A $(1-\alpha)\,100\%$ confidence interval for the population variance σ^2
(where the population is assumed normal)

$$\left(\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2}},\ \frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2}} \right)$$

$n-1$ degrees of freedom



A $(1-\alpha)\,100\%$ confidence interval for the population standard deviation σ (where the population is assumed normal)

$$\left(\sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2}}},\ \sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2}}} \right) \quad \text{OR} \quad \left(s\cdot\sqrt{\frac{(n-1)}{\chi^2_{\alpha/2}}},\ s\cdot\sqrt{\frac{(n-1)}{\chi^2_{1-\alpha/2}}} \right)$$

$n-1$ degrees of freedom

A machine makes ½-inch ball bearings. In a random sample of 41 bearings, the sample standard deviation of the diameters of the bearings was 0.02 inch. Assume that the diameters of the bearings are approximately normally distributed. Construct a 90% confidence interval for the standard deviation of the diameters of the bearings.

$$\left(\sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2}}}, \sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2}}} \right)$$

The following random sample was obtained from $N(\mu, \sigma^2)$ distribution:

16	12	18	13	21	15	8	17

$\bar{x} = 15, \qquad s^2 = 16, \qquad s = 4.$

→ Construct a 95% confidence interval for the overall standard deviation.

$$\left(\sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2}}}, \sqrt{\frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2}}} \right)$$

→ Construct a 95% confidence lower bound for the overall standard deviation.

→ Construct a 95% confidence upper bound for the overall standard deviation.

Confidence Intervals for Difference of Proportions

Consider two dichotomous populations, with "success" proportions p_1 and p_2, respectively. Consider the sample proportions

$$\hat{p}_1 = \frac{x_1}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{x_2}{n_2}$$

where n_1 and n_2 are the sample sizes and x_1 and x_2 are the numbers of "successes" in the two samples from populations 1 and 2, respectively.

If n_1 and n_2 are large, then $(\hat{p}_1 - \hat{p}_2)$ is approximately normal with mean $p_1 - p_2$ and standard deviation $\sqrt{\dfrac{p_1 \cdot (1-p_1)}{n_1} + \dfrac{p_2 \cdot (1-p_2)}{n_2}}$.

The confidence interval for the difference between two population proportions $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \cdot (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1 - \hat{p}_2)}{n_2}}$$

In a comparative study of two new drugs, A and B, 120 patients were treated with drug A and 150 patients with drug B, and the following results were obtained.

	Drug A	Drug B
Cured	78	111
Not cured	42	39
Total	120	150

Construct a 95% confidence interval for the difference in the cure rates of the two drugs.

Confidence Intervals for Difference of Means

Two **independent** samples:

$X_1, X_2, \ldots, X_{n_1}$

from population 1

mean μ_1, std. dev. σ_1

$Y_1, Y_2, \ldots, Y_{n_2}$

from population 2

mean μ_2, std. dev. σ_2

Two **independent** samples:

$X_1, X_2, \ldots, X_{n_1}$ from population 1, mean μ_1, std. dev. σ_1

$Y_1, Y_2, \ldots, Y_{n_2}$ from population 2, mean μ_2, std. dev. σ_2

If we can assume that population 1 and population 2 are both normal, and the standard deviations are of the two populations are equal (i.e., $\sigma_1 = \sigma_2 = \sigma$), then we can use

$$\left(\overline{X} - \overline{Y}\right) \pm t_{\alpha/2} \cdot s_{\text{pooled}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where

$$s_{\text{pooled}}^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}.$$

The number of degrees of freedom $= n_1 + n_2 - 2$.

A national equal employment opportunities committee is conducting an investigation to determine if female employees are as well paid as their male counterparts in comparable jobs. Random samples of 14 males and 11 females in junior academic positions are selected, and the following calculations are obtained from their salary data.

	Male	Female
Sample Mean	$48,530	$47,620
Sample Standard deviation	780	750

Assume that the populations are normally distributed with equal variances.

Construct a 95% confidence interval for the difference between the mean salaries of males and females in junior academic positions.


That's all Folks!