Ch. 8: Confidence Intervals

8.1 Estimating Population Means (σ known)

Recall, inferential statistics is in the business of providing estimates for population parameters (e.g. mean, standard deviation, proportion).

- If a estimate is a single value, its called a point estimate
  e.g. sample mean $\bar{x}$ is point estimate of population mean μ
- point estimate is unbiased if it does not consistently underestimate or overestimate a population parameter.

Sometimes, it is more appropriate to estimate a range of values for a population parameter.

Def: an interval estimate is a range of possible values of a population parameter.

Level of confidence, denoted c, is percentage of all samples of given size that produce interval estimates that contain the actual parameter.

Confidence interval is interval estimate associated w/ level of confidence.

Margin of error

Example college student researching study habits found $\bar{x} = 15.7$ hours/week. If margin of error is $E = 0.6$ hours, construct a 95% confidence interval.

$(\bar{x} - E, \bar{x} + E)$ $\quad$ $(15.1, 16.3)$

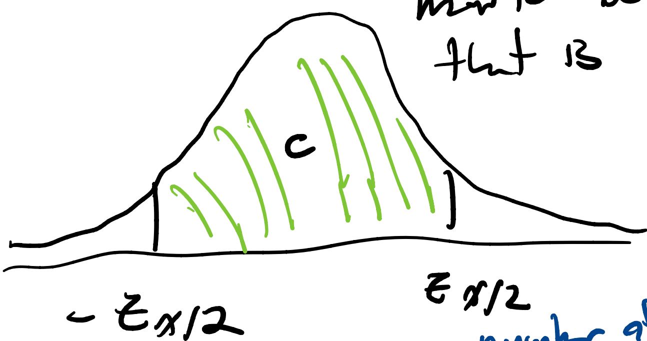
Using Standard Normal to Estimate the Population Mean

Requirements

1. All samples "=" probability of being chosen
2. population standard deviation σ is known
3. $n \geq 30$ <u>or</u> normal

Critical z-values $-z_{\alpha/2}$, $z_{\alpha/2}$ mark boundaries for area that is level of confidence.

$$\alpha = 1 - c$$



Margin of error, $E = (z_{\alpha/2}) \cdot (\sigma_{\bar{x}})$

number of standard deviations for given confidence

standard dev. of sampling distribution

$$= (z_{\alpha/2}) \cdot \frac{\sigma}{\sqrt{n}}$$

where n is sample size, σ population std.

Finding critical z-values

c	$\alpha = 1 - c$	$z_{\alpha/2}$
0.80	0.20	1.28
0.90	0.10	1.44
0.95	0.05	1.96
0.98	0.02	2.33
0.99	0.01	2.576

Confidence interval for sample mean:
$(\bar{x} - E, \bar{x} + E)$ where $\bar{x}$ is the sample mean.

<u>Example</u> Volunteers contact $n = 35$ mental health hotlines across nation to find $\bar{x} = 35$ calls/month. Construct a 95% confidence interval for mean number of calls per month. Assume $\sigma = 6.5$ calls/month.

Step 1: find point estimate

Step 2: find margin of error (MoE)

Step 3: Find endpoints

Example: $n = 126$ seeds, $\bar{x} = 13.9$ days

Assuming $\sigma = 1.4$ days, create 90% confidence interval for number of days it takes hydrangea seeds to germinate.

Minimum Sample Size

Increase n $\Rightarrow$ increase level of confidence = reduce MoE

Question: what is the fewest number of samples needed to obtain desired MoE?

Minimum sample size

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

where σ is pop. std., E is desired M.E.

Example: what is minimum sample size needed to be 90% confident that sample mean within 2 units of population mean?

§8.2 Student's t-Distribution

In real life, the standard deviation σ is unknown.

New distribution allows you to calculate estimates of $\bar{x}$

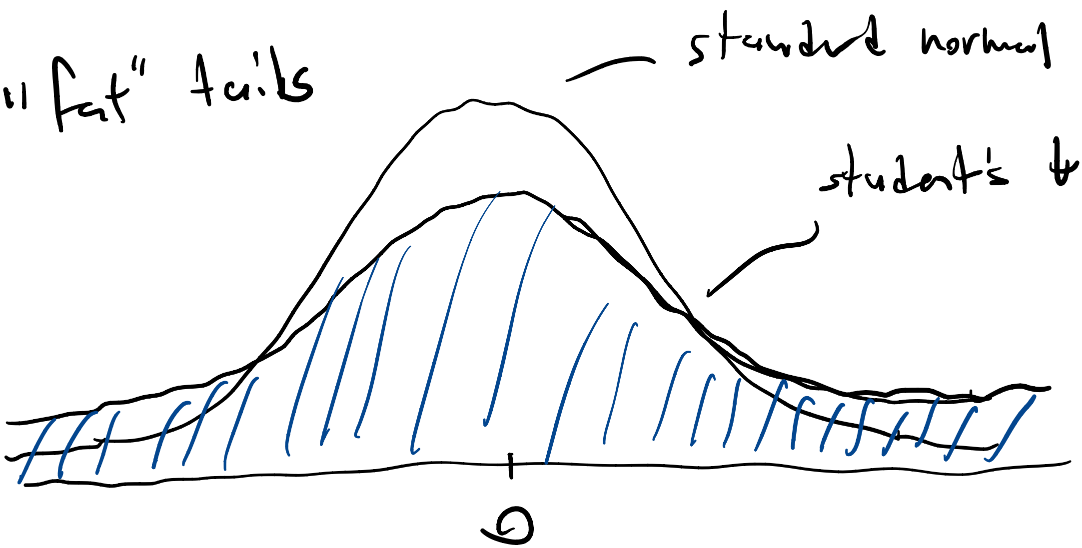
when σ is unknown.

Can use t-distribution if pop. normally distributed or sufficiently large.

t-distribution has parameter

df = "degrees of freedom"

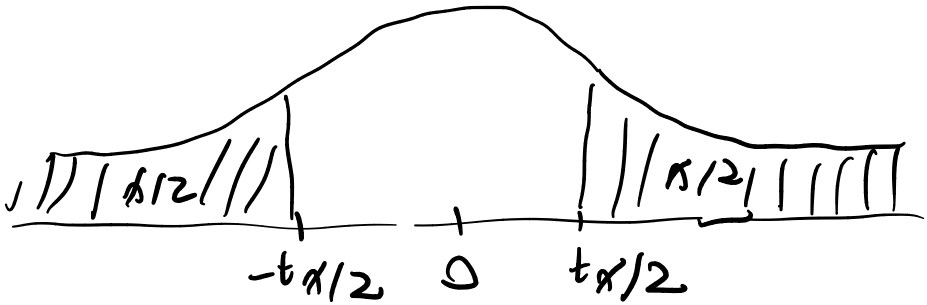
As df is large $\longrightarrow$ standard normal



More df $\Rightarrow$ less area under tails

Less df $\Rightarrow$ more area under tails

Finding t_α where $\alpha = 1 - c$



Can look up in a table
(usually looking for critical t-values)

8.3 Estimating Population Means σ unknown

Requirements
- all samples equal probability
- σ unknown
- $n \geq 30$ or approx. normal dist.

MoE for σ-unknown given by
$$E = (t_{\alpha/2})\left(\frac{s}{\sqrt{n}}\right)$$
where s is sample std, $t_{\alpha/2}$ is critical value for confidence $c = 1-\alpha$ such that area under the t-dist. w/ $n-1$ degrees of freedom to right of $t_{\alpha/2}$ is equal to $\frac{\alpha}{2}$

(compare to $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$)

Now, construct confidence int. as before $(\bar{x} - E, \bar{x} + E)$

<u>Example</u> Marketing agency has sample of $n = 756$ cars w/ $\bar{x} = \$27{,}409$
Construct 95% confidence interval for mean price of cars sold in the area.

<u>Example</u> Find 99% confidence for $\bar{x}$

Household Water Used for Brushing Teeth (in Gallons per Day)

0.485	0.428	0.390	0.308	0.231
0.587	0.516	0.465	0.370	0.282
0.412	0.367	0.336	0.269	0.198
0.942	0.943	0.940	0.941	0.946
0.868	0.898	0.889	0.910	0.927
0.925	0.950	0.959	0.948	0.956
0.805	0.810	0.839	0.860	0.861
0.515	0.463	0.420	0.326	0.243

<u>Ans</u>: $(0.5240, 0.7624)$

8.4 Estimating Population Proportions

Another parameter comonly estimated besides μ is p. The pt. estimate for p is $\hat{p} = \frac{x}{n}$

Requirements for constructing a CI for pop. proportion:

1. All samples of given size equal probability of being chosen
2. Conditions for binomial are met
   - fixed num. of identical trials
   - trials independent
   - 2 outcomes
3. $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$

MoE of CI for pop. proportion:

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$ where

- $z_{\alpha/2}$ is critical value for level of confidence $c = 1 - \alpha$ (i.e. area under standard normal to right of $z_{\alpha/2}$ is $\alpha/2$)

- n is the sample size

The CI constructed as before:

$$(\hat{p} - E, \hat{p} + E)$$

Example: Marist college poll has sample of $n = 1401$ likely voters with 52.5% saying they are voting for Harris and 47.5% saying voting for Trump. Construct a 99% confidence interval.

Say Harris is "success" so $\hat{p} = 0.525$

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

$\alpha = 1 - 0.99 = 0.01$ $\quad z_{\alpha/2} = 2.576$

$$E = 2.575 \cdot \sqrt{\frac{0.525 \cdot 0.475}{1401}}$$

$$= 0.024$$

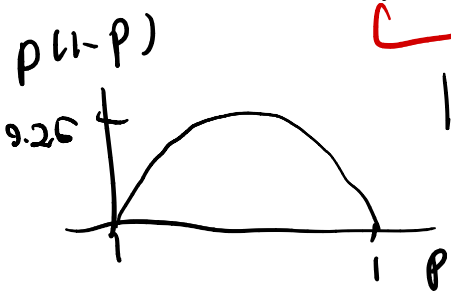
Hence, the MoE is 2.4%.

So a 99% confidence interval for proportion of voters voting for Harris is $(\hat{p} - E, \hat{p} + E)$, i.e.

$$(0.501, 0.549)$$

We also can estimate minimum sample size at a given level of confidence:

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



Hence, the closer the race, the greater the sample size needed.

8.5 Estimating Population Variances

Point estimate of σ is s

" " of σ^2 is s^2

Requirements

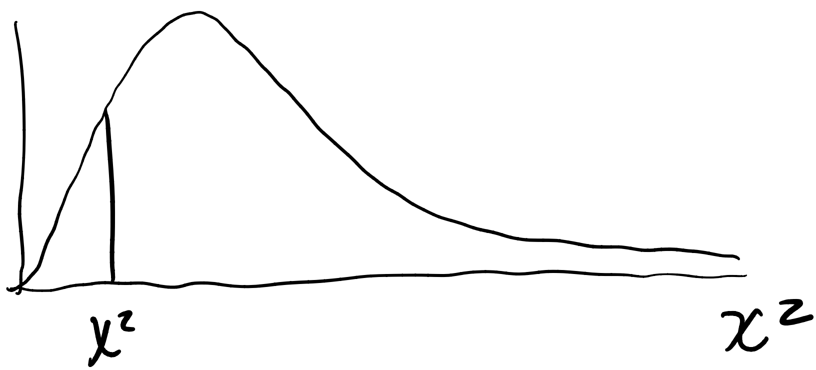
- all samples of given size equal probability of being chosen
- population distribution approximately normal

To estimate population variance, will need new distribution called χ^2-dist.

χ^2 completely determined by df

$\mu = df$

$\sigma^2 = 2\,df$



Confidence interval for pop. variance

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

s^2 is sample variance

$\chi^2_{\alpha/2}$, $\chi^2_{(1-\alpha/2)}$ critical values

for level of confidence $c = 1-\alpha$

CI for pop. std.

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}}$$

Example: comercial bakery testing variance of weights of cookies. Random sample of $n=15$ cookies $s^2 = 3.4$. Build a 95% CI for variance of weights of all cookies.