Agenda

- Exam on Monday April 30.
- Practice questions posted.
- Friday’s class is for review (no studio)

Today

- Review of critical values and quantiles.
- Computing $z$, $t$, $\chi^2$ confidence intervals for normal data.
- Conceptual view of confidence intervals.
- Confidence intervals for polling (Bernoulli distributions).
Review of critical values and quantiles

- **Quantile**: left tail $P(X < q_\alpha) = \alpha$
- **Critical value**: right tail $P(X > c_\alpha) = \alpha$

Letters for critical values:

- $z_\alpha$ for $N(0, 1)$
- $t_\alpha$ for $t(n)$
- $c_\alpha$, $x_\alpha$ all purpose

$q_\alpha$ and $z_\alpha$ for the standard normal distribution.
Concept question

1. $z_{0.025} =$
   
   (a) -1.96  (b) -0.95  (c) 0.95  (d) 1.96  (e) 2.87
Concept question

1. \( z_{025} = \)
   
   (a) -1.96   (b) -0.95   (c) 0.95   (d) 1.96   (e) 2.87

2. \( -z_{16} = \)
   
   (a) -1.33   (b) -0.99   (c) 0.99   (d) 1.33   (e) 3.52
Computing confidence intervals from normal data

Suppose the data $x_1, \ldots, x_n$ is drawn from $\text{N}(\mu, \sigma^2)$
Confidence level $= 1 - \alpha$

- **$z$ confidence interval for the mean ($\sigma$ known)**

\[
\left[ \bar{x} - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \quad \bar{x} + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} \right]
\]

- **$t$ confidence interval for the mean ($\sigma$ unknown)**

\[
\left[ \bar{x} - \frac{t_{\alpha/2} \cdot s}{\sqrt{n}}, \quad \bar{x} + \frac{t_{\alpha/2} \cdot s}{\sqrt{n}} \right]
\]

- **$\chi^2$ confidence interval for $\sigma^2$**

\[
\left[ \frac{n-1}{c_{\alpha/2}} s^2, \quad \frac{n-1}{c_{1-\alpha/2}} s^2 \right]
\]

- $t$ and $\chi^2$ have $n - 1$ degrees of freedom.
Suppose \( x_1, \ldots, x_n \sim N(\mu, \sigma^2) \) with \( \sigma \) known.

The rule-of-thumb 95% confidence interval for \( \mu \) is:

\[
\left[ \bar{x} - 2 \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 2 \frac{\sigma}{\sqrt{n}} \right]
\]

A more precise 95% confidence interval for \( \mu \) is:

\[
\left[ \bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right]
\]
The data 4, 1, 2, 3 is drawn from $N(\mu, \sigma^2)$ with $\mu$ unknown.

1. Find a 90% $z$ confidence interval for $\mu$, given that $\sigma = 2$.

For the remaining parts, suppose $\sigma$ is unknown.

2. Find a 90% $t$ confidence interval for $\mu$.

3. Find a 90% $\chi^2$ confidence interval for $\sigma^2$.

4. Find a 90% $\chi^2$ confidence interval for $\sigma$.

5. Given a normal sample with $n = 100$, $\bar{x} = 12$, and $s = 5$, find the rule-of-thumb 95% confidence interval for $\mu$. 
Conceptual view of confidence intervals

- Computed from data $\Rightarrow$ **interval statistic**
- ‘Estimates’ a parameter of interest $\Rightarrow$ **interval estimate**
- Width = measure of precision
- Confidence level = measure of performance
- Confidence intervals are a frequentist method.
  - No need for a prior, only uses likelihood.
  - Frequentists **don’t assign probabilities to hypotheses**
  - A 95% confidence interval of $[1.2, 3.4]$ for $\mu$ **doesn’t mean** that $P(1.2 \leq \mu \leq 3.4) = 0.95$.
- We will compare with Bayesian probability intervals later.

Applet:
http://mathlets.org/mathlets/confidence-intervals/
The quantities $n$, $c = \text{confidence}$, $\bar{x}$, $\sigma$ all appear in the $z$ confidence interval for the mean.

How does the width of a confidence interval for the mean change if:

1. we increase $n$ and leave the others unchanged?
2. we increase $c$ and leave the others unchanged?
3. we increase $\mu$ and leave the others unchanged?
4. we increase $\sigma$ and leave the others unchanged?

(A) it gets wider  (B) it gets narrower  (C) it stays the same.
Intervals and pivoting

$\bar{x}$: sample mean (statistic)

$\mu_0$: hypothesized mean (not known)

**Pivoting:** $\bar{x}$ is in the interval $\mu_0 \pm 2.3 \iff \mu_0$ is in the interval $\bar{x} \pm 2.3$.

```
-2 -1 0 1 2 3 4

\mu_0 ± 1
\bar{x} ± 1
\mu_0 ± 2.3
\bar{x} ± 2.3
```

**Algebra of pivoting:**

$$\mu_0 - 2.3 < \bar{x} < \mu_0 + 2.3 \iff \bar{x} + 2.3 > \mu_0 > \bar{x} - 2.3.$$
Suppose $x_1, \ldots, x_n \sim \mathcal{N}(\mu, \sigma^2)$ with $\sigma$ known.

Consider two intervals:

1. The $z$ confidence interval around $\bar{x}$ at confidence level $1 - \alpha$.
2. The $z$ non-rejection region for $H_0 : \mu = \mu_0$ at significance level $\alpha$.

Compute and sketch these intervals to show that:

$$\mu_0 \text{ is in the first interval} \iff \bar{x} \text{ is in the second interval.}$$
Solution

Confidence interval: \( \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \)

Non-rejection region: \( \mu_0 \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \)

Since the intervals are the same width they either both contain the other’s center or neither one does.
Polling: a binomial proportion confidence interval

Data \( x_1, \ldots, x_n \) from a Bernoulli(\( \theta \)) distribution with \( \theta \) unknown. A conservative normal\(^\dagger\) \((1 - \alpha)\) confidence interval for \( \theta \) is given by

\[
\left[ \bar{x} - \frac{z_{\alpha/2}}{2\sqrt{n}}, \bar{x} + \frac{z_{\alpha/2}}{2\sqrt{n}} \right].
\]

Proof uses the CLT and the observation \( \sigma = \sqrt{\theta(1 - \theta)} \leq 1/2 \).

Political polls often give a margin-of-error of \( \pm 1/\sqrt{n} \). This rule-of-thumb corresponds to a 95% confidence interval:

\[
\left[ \bar{x} - \frac{1}{\sqrt{n}}, \bar{x} + \frac{1}{\sqrt{n}} \right].
\]

(The proof is in the class 21 notes.)

Conversely, a margin of error of \( \pm 0.05 \) means 400 people were polled.

\(^\dagger\)There are many types of binomial proportion confidence intervals.

Board question

For a poll to find the proportion \( \theta \) of people supporting X we know that a \((1 - \alpha)\) confidence interval for \( \theta \) is given by

\[
\left[ \bar{x} - \frac{z_{\alpha/2}}{2\sqrt{n}}, \; \bar{x} + \frac{z_{\alpha/2}}{2\sqrt{n}} \right].
\]

1. How many people would you have to poll to have a margin of error of .01 with 95% confidence? (You can do this in your head.)

2. How many people would you have to poll to have a margin of error of .01 with 80% confidence. (You’ll want R or other calculator here.)

3. If \( n = 900 \), compute the 95% and 80% confidence intervals for \( \theta \).