8.1 – Confidence Intervals

Generic large sample confidence intervals:

Def'n: A confidence interval (CI) for a parameter θ is an interval estimate of plausible values for θ. With a chosen degree of confidence, the CI's construction is such that the value of θ is captured between the statistics L and U, the lower and upper endpoints of the interval, respectively.

The confidence level of a CI estimate is the success rate of the *method* used to construct the interval (as opposed to confidence in any particular interval). The generic notation is $100(1-\alpha)\%$. Typical values are 90%, 95%, and 99%.

Ex8.1) Using 95% and the upcoming method to construct a CI, the method is "successful" 95% of the time. That is, if this method was used to generate an interval estimate over and over again with different samples, in the long run, 95% of the resulting intervals would capture the true value of θ.

Many large sample CIs have the form:

estimator ± (critical value) × (standard error)

where "estimator" is a statistic $\hat{\theta}$ used to estimate parameter θ,

"standard error" is a statistic $\hat{\sigma}_{\hat{\theta}}$ used to estimate std. dev. of estimator $\hat{\theta}$,

"critical value" is a fixed number z such that if Z has a std. norm. dist'n, then $P(-z \le Z \le z) = 1 - \alpha$ = confidence level

The product of the "standard error" and "critical value" is the *margin of error*.

Note: critical value z often denoted by $z_{\alpha/2}$, where the notation reflects $P(Z > z) = \alpha/2$.

Ex8.2) If the confidence level is 95%, what is the critical value?

Table 8X0 – Critical values for usual confidence levels

100(1 – α)%	**α**	**$\alpha/2$**	**$z_{\alpha/2}$**
90%	0.10	0.050	1.645
95%	0.05	0.025	1.960
99%	0.01	0.005	2.576

The estimator $\hat{\theta}$ and its (estimated) standard error $\hat{\sigma}_{\hat{\theta}}$ are defined so that, when the sample size n is sufficiently large, the sampling distribution of

$$\frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}} \mathrel{\dot\sim} N(0,1)$$

Consequently,

$$P\left(-z \le \frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}} \le z\right) \approx 1-\alpha$$

Algebraic manipulation yields

$$P\left(\hat{\theta} - z\hat{\sigma}_{\hat{\theta}} \le \theta \le \hat{\theta} + z\hat{\sigma}_{\hat{\theta}}\right) \approx 1-\alpha$$

Confidence Interval for a Population Mean
A $(1 - \alpha)100\%$ CI for μ is

$$\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Assumptions: random sample, $n \geq 30$ OR the population is normal, σ is known

Ex8.3) Professor Frink wants to determine the mean radius of all of Jupiter's several moons. From 4 moons, he discovers a sample mean of 2103.75 km. Suppose $\sigma = 816$ km. Construct and interpret a 90% confidence interval for the population mean radius of Jupiter's moons.

Direct interpretation:

Never write $P(\bar{x}_L \leq \mu \leq \bar{x}_U) = 0.90$. Wrong conceptual interpretation.
Correct conceptual interpretation: If many samples were obtained and corresponding intervals calculated, about 90% of the intervals would cover μ.

The *margin of error* for a CI:
1. Increases as the confidence level increases; 2. Decreases as the sample size increases.

Ex8.4) Starting from Ex8.3),

a) If confidence level changes to 99%, interval is

b) Suppose $n = 8$, then the standard error =

and interval is

Confidence Interval for a Population Mean, σ unknown
This situation is fairly common, so the sample data must estimate σ.
Recall $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$. We now have $Z \neq \dfrac{\bar{X} - \mu}{s / \sqrt{n}} = t$, where t is a diff. standardized variable.
The value of s may not be all that close to σ, especially when n is small. Consequently, there is extra variability and the distribution of t is more spread out than the z curve.

t-distributions (see Section 7.3, p. 311):
As with normal curves, there exists a family of t-curves.
The normal distribution has 2 parameters: μ and σ.
The t-distribution has a single parameter: *degrees of freedom* (*df*).
Range of t: similar to range of z; range of *df*: 1, 2, 3,…, ∞

Properties of the t-distribution:
1. The t-curve, with any fixed *df*, is bell-shaped and centered at 0 (just like the z-curve).
2. Each t-curve is more spread out than the z-curve. $t_{\alpha/2} > z_{\alpha/2}$
3. As *df* increases, the spread of corresponding t-curve decreases.
4. As *df* increases, the corresponding sequence of t-curves approaches the z-curve.

One-Sample CI, σ unknown
A $(1 - \alpha)100\%$ CI for μ is

$$\bar{x} \pm t_{n-1.\,\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Assumptions: random sample, $n \geq 30$ OR the population is normal, σ is unknown

Table III gives critical values appropriate for each of the 90%, 95%, 99% confidence levels, and several other confidence levels.

Ex8.5) A naïve astrophysicist wants to determine the mean radius of all of Jupiter's numerous moons. Only taking the 4 largest moons, he discovers a sample mean of 2103.75 km with a standard deviation of 495 km. Construct a 90% confidence interval for the population mean radius of Jupiter's moons. Assume radii follow a normal distribution. (Note: changes were made in class to this example to make the assumptions hold.)

One-sided confidence "intervals"
These differ from "two-sided" ones. A $100(1 - \alpha)\%$ upper-confidence interval for μ is

$$-\infty \leq \mu \leq u = \bar{x} + z_{\alpha}\left(\frac{\sigma}{\sqrt{n}}\right) \qquad \left[\text{OR } -\infty \leq \mu \leq u = \bar{x} + t_{\alpha}\left(\frac{s}{\sqrt{n}}\right)\right]$$

The $100(1 - \alpha)\%$ lower-confidence interval for μ has similar computation.

Choosing the sample size n:

Consider the CI as $\bar{x} \pm E$, where the margin of error: $E \approx z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

Differing from textbook (which uses $L = 2E$), solving for n here gives

$$n \approx \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = 4\left(\frac{z_{\alpha/2}\sigma}{L_0}\right)^2$$

Round up n to next integer. Replace σ by a prior estimate or, for a population that is not too skewed, use $\sigma \approx$ (range)/6 since 99.7% of the data is approximately within $3s$ of the sample mean. (This situation is not common.)

Ex8.6) You wish to estimate the mean cost of textbooks so that the margin of error is within \$10 of the true population mean. Most textbook prices are between \$70 and \$220. How large of a sample size is required to be 99% confident in achieving the mentioned level of accuracy? Assume textbook cost is normally distributed.

8.2 Hypothesis Testing

Def'n: A null hypothesis is a claim about a population parameter that is assumed to be true until it is declared false.

An alternative hypothesis is a claim about a population parameter that will be true if the null hypothesis is false.

In carrying out a test of H_0 vs. H_A, the hypothesis H_0 is "rejected" in favour of H_A only if sample evidence strongly suggests that H_0 is false. If the sample does not contain such evidence, H_0 is "not rejected" or you "fail to reject" it.

NEVER "accept" H_0 or H_A…for different reasons.

Ex8.7) H_0: $\mu = 2.8$ $\qquad$ H_A: $\mu \neq 2.8$

$\uparrow$ population parameter $\qquad$ $\uparrow$ hypothesized value or "claim"

Def'n: A two-sided test has "rejection regions" in both tails.

A one-sided test has a "rejection region" in one tail.

Ex8.8)

a) H_0: $\mu = 15$ $\qquad$ H_A: $\mu = 15$

b) H_0: $\mu = 123$ $\qquad$ H_A: $\mu = 125$

c) H_0: $\mu = 123$ $\qquad$ H_A: $\mu < 123$

d) H_0: $\mu \geq 123$ $\qquad$ H_A: $\mu < 123$

e) H_0: $p = 0.4$ $\qquad$ H_A: $p > 0.6$

f) H_0: $p = 1.5$ $\qquad$ H_A: $p > 1.5$

g) H_0: $\hat{p} = 0.1$ $\qquad$ H_A: $\hat{p} \neq 0.1$

	Two-Sided Test	**One-Sided Test (Lower)**	**One-Sided Test (Upper)**
Sign for H_0	$=$	$=$ or $\geq$	$=$ or $\leq$
Sign for H_A	$\neq$	$<$	$>$
"Rejection region"	In both tails	In the lower tail	In the upper tail

Ex8.9)

Is the mean different than μ_0?

Is the mean lower than μ_0?

Is the mean lower or still the same than μ_0?

Is the mean higher than μ_0?

p-values

A test statistic is the function of the sample data on which a conclusion to reject or fail to reject H_0 is based. For example, Z and t are test statistics.

The p-value is a measure of plausibility between the hypothesized value for a population characteristic and the observed sample. Keep in mind that we usually prefer to be inconsistent with H_0 so we can reject it. **The smaller the *p*-value, the more likely we reject H_0.**

Using the "judgment approach" for rejection,

$0.01 > p$-value > 0 → strong to convincing evidence against H_0
$0.05 > p$-value > 0.01 → moderate to strong evidence against H_0
$0.10 > p$-value > 0.05 → suggestive to moderate evidence against H_0, yet inconclusive
$1 > p$-value > 0.1 → weak evidence against H_0

The "significance level approach":

$$\text{reject } H_0 \text{ if } p\text{-value} \leq \alpha$$
$$\text{do not reject } H_0 \text{ if } p\text{-value} > \alpha$$

Common choices for α are 0.01, 0.05 and 0.1, depending on the nature of the test.

PROBLEMS:

a) If you're comparing to α and do not mention the *p*-value, do you know if the *p*-value is 0.045 or 0.000 001?

b) If we use a "cut-off" like $\alpha = 0.05$, does it make sense to conclude differently between *p*-values of 0.049 and 0.051?

Solution: ALWAYS report your *p*-value! They are far more informative for a test than comparison to a significance level. Also, the "judgment approach" is preferred for decisions.

Errors in Hypothesis Testing
In any hypothesis test, there is 1 of 2 choices: reject or not reject. There is also 1 of 2 choices as the test applies to reality: H_0 is true or H_0 is false.

		Actual situation	
		H_0 is true	H_0 is false
Decision	Do not reject H_0	Correct Decision	Type II or β error
	Reject H_0	Type I or α error	Correct Decision

A Type I error occurs when a true null hypothesis is rejected. The value of α represents the prob. of committing this type of error; that is,

$$\alpha = P(H_0 \text{ is rejected} \mid H_0 \text{ is true})$$

The value of α represents the *significance level* of the test.

A Type II error occurs when a false null hypothesis is not rejected. The value of β represents the prob. of committing a Type II error; that is,

$$\beta = P(H_0 \text{ is not rejected} \mid H_0 \text{ is false})$$

The value of $1 - \beta$ is called the *power of the test*. It represents the probability of NOT making a Type II error. Or, power = P(rejecting H_0 | H_0 is false).

Ex8.10) H_0: "innocent until proven guilty"

		Actual situation	
		Innocent	Guilty
Jury's decision	Find not guilty	Correct Decision	Type II or β error
	Find guilty	Type I or α error	Correct Decision

Note:
1. These 2 errors are dependent. For a fixed n, lowering α will raise β and vice versa.
2. Decreasing α and β simultaneously requires increasing the sample size.
3. When H_0 is false, β increases as the true value approaches the "claim". Conversely, β decreases as the true value moves away from the "claim".

Relating Hypothesis Tests and CI
If $[L, U]$ is a $100(1 - \alpha)\%$ CI for parameter θ, the test of size α of the hypothesis

$$H_0: \theta = \theta_0 \qquad H_A: \theta \neq \theta_0$$

will lead to rejection of H_0 if and only if θ_0 is NOT in $100(1 - \alpha)\%$ CI.

- The CI provides a range of likely values for a parameter at a stated conf. level.
- A lower-tailed test associates with an upper confidence bound (and vice versa).

Steps of a Hypothesis Test
1. Specify variable/parameter.
2. State the H_0 and H_A. (Optional: Select α for the test.)
3. Check assumptions. Determine appropriate test statistic.
4. Compute sample statistics, then test statistic.
5. Determine exact p-value (or range).
6. Make a decision and conclude within the context of the problem.

Test for Population Mean, σ Known
Assumptions: random sample, $n \geq 30$ OR the population is normal, σ is known

H_0: $\mu = \mu_0$ H_A: $\mu \neq \mu_0$

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$ → find p-value from $N(0, 1)$

Ex8.11) A website claims the average length of an episode is 21.104 min. Suppose we know that σ = 0.364 min. A bored student takes a random sample of 7 episodes and found their mean length to be 21.16 min. It is known that the length of all episodes has an approx. normal distribution. Test the above claim using both approaches and α = 0.01.

The sample is random, σ is known, n is small but the population is normally distributed.
→ Thus, we may use the z-distribution.

H_0: $\mu = 21.104$ H_A: $\mu \neq 21.104$

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{21.16 - 21.104}{0.364 / \sqrt{7}} = 0.41$$

p-value $= 2P(Z > 0.41) = 2(1 - P(Z < 0.41)) = 2(1 - 0.6591) = 0.6818$

Using the "judgment approach", we have weak evidence against H_0. Using the "significance level approach" and α = 0.01, we do not reject H_0. Either way, we conclude that the website's claim may be valid.

(Note that one-tailed tests similar to this example were discussed in class.)

Test for Population Mean, σ Unknown
Assumptions: random sample, $n \geq 30$ OR the population is normal, σ is unknown

H_0: $\mu = \mu_0$ H_A: $\mu \neq \mu_0$

$$T_0 = \frac{\overline{X} - \mu_0}{S / \sqrt{n}}$$ → find p-value from t_{n-1}

Ex8.12) A website claims the average length of an episode is 21.104 min. Another bored student takes a random sample of 7 episodes and found their mean length to be 21.16 min with a standard deviation of 0.20 min. It is known that the length of all episodes has an approx. normal distribution. Test the above claim using both approaches and $\alpha = 0.01$.

The sample is random; n is small, yet the population is normally distributed; σ is unknown. → Thus, we may use a t-distribution; specifically, t_6.

H_0: $\mu = 21.104$ $\qquad$ H_A: $\mu \neq 21.104$

$$T_0 = \frac{\overline{X} - \mu_0}{S / \sqrt{n}} = \frac{21.16 - 21.104}{0.20 / \sqrt{7}} = 0.741 \sim t_6$$

$0 \quad < T_0 = 0.741 < 1.440$
$0.5 > \qquad\qquad > 0.10$ → but test is two-tailed, so multiply range by two
$1 \quad > \quad p\text{-value} \quad > 0.20$ → actual p-value range is $(0.20, 1)$

In fact, Excel's TDIST() function produces an exact p-value of 0.4867.

Using the "judgment approach", we have weak evidence against H_0. Using the "significance level approach" and $\alpha = 0.01$, we do not reject H_0. Either way, we conclude that the website's claim may be valid.

(Note that one-tailed tests similar to this example were discussed in class.)
(Small examples to give more practice calculating p-value ranges also occurred.)