Unit 4 · Statistical Inference Cheat Sheet

Unit 4 focus: confidence intervals for a proportion and a mean, conditions and interpretation

1 KEY VOCABULARY

point estimate	A statistic used to estimate a parameter.	$\hat{p} = 0.6$ estimates p
standard error	Estimated SD of the statistic.	$\sqrt{\hat{p}(1-\hat{p})/n}$
margin of error	Critical value × standard error.	± 0.048
confidence level	Success rate of the method in repeated samples.	95%
critical value	z^* or t^* for the confidence level.	$z^* = 1.96$

2 BUILDING AN INTERVAL

Piece	What to use
Form	statistic ± (critical value)(standard error)
Conditions	**Random**; **10%**: $n \le 0.1N$; **Large counts** $n\hat{p}, n(1-\hat{p}) \ge 10$ or **Normal** ($n \ge 30$ or no strong skew)
z^*	90%: 1.645 · 95%: 1.96 · 99%: 2.576

Higher confidence $\Longrightarrow$ wider. Larger $n \Longrightarrow$ narrower (ME shrinks like $1/\sqrt{n}$).

3 ONE-PROPORTION Z-INTERVAL

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Example: 240 of 400 random voters support a bond. 95% CI?

$\hat{p}$: $\frac{240}{400} = 0.60$; counts 240 and 160 ≥ 10

SE: $\sqrt{0.6(0.4)/400} \approx 0.0245$

ME: $1.96(0.0245) \approx 0.048$

Answer: (0.552, 0.648)

4 ONE-MEAN T-INTERVAL

$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$ with $df = n - 1$.

Example: $n = 25$, $\bar{x} = 82$, $s = 10$. 95% CI?

t^*: df = 24, $t^* = 2.064$

SE: $10/\sqrt{25} = 2$

ME: $2.064(2) = 4.128$

Answer: (77.87, 86.13)

5 INTERPRETING AND PLANNING

Interval	"We are 95% confident that the interval from 0.552 to 0.648 captures the true proportion of *all voters* who support the bond."
Level	"If we took many samples, about 95% of the intervals would capture the true proportion."
Sample size	ME ≤ 0.03 at 95%: $n \ge (\frac{1.96}{0.03})^2(0.5)(0.5) = 1067.1$, so $n = 1068$

6 QUICK REMINDERS & COMMON MISTAKES

✔ Do remember

- Check conditions before computing.
- Intervals estimate a **parameter**, never a statistic.
- Use $\hat{p} = 0.5$ for sample size when you have no guess.

✘ Common mistakes

- "95% chance p is in this interval": the method captures p 95% of the time
- Using z^* for a mean with s: use t^* with $df = n - 1$
- Rounding n down: always round up

Unit 4 · Statistical Inference Cheat Sheet

More practice: significance tests, p-values, Type I and II errors, power and two-sample procedures

1 THE FOUR-STEP PROCESS

State	H_0 (no effect, "="), H_a ($<$, $>$ or $\neq$), α; define the parameter
Plan	name the test; check Random, 10%, Large counts / Normal
Do	test statistic, then the p-value
Conclude	compare p to α and answer in context

Hypotheses are always about **parameters** (p, μ), never $\hat{p}$ or $\bar{x}$.

2 TEST STATISTICS

Test	Statistic
One proportion	$z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$
One mean	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$, $df = n - 1$
Two proportions	$z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_C(1-\hat{p}_C)(\frac{1}{n_1} + \frac{1}{n_2})}}$

$$\text{statistic} = \frac{\text{observed} - \text{hypothesized}}{\text{SD of statistic}}$$

3 ONE-PROPORTION Z-TEST

$H_0 : p = 0.5$, $H_a : p > 0.5$, $\alpha = 0.05$.

Example: 270 of 500 randomly chosen students prefer the new schedule.

$\hat{p}$: $\frac{270}{500} = 0.54$

z: $\frac{0.54 - 0.5}{\sqrt{0.25/500}} \approx 1.79$

p-value: $P(Z > 1.79) \approx 0.037 < 0.05$

Answer: reject H_0: convincing evidence $p > 0.5$

4 P-VALUES AND CONCLUSIONS

p-value: the probability of a result this extreme *if H_0 is true*	**small p = evidence for H_a**
$p < \alpha$	**reject H_0; convincing evidence for H_a**
$p \geq \alpha$	**fail to reject H_0; not convincing evidence**
Never write	**"accept H_0" or "proves H_0"**

5 ERRORS AND POWER

	H_0 **true**	H_0 **false**
Reject H_0	**Type I** (prob. α)	✔ correct (power)
Fail to reject	✔ correct	**Type II** (prob. β)

Power = $1 - \beta$. Power rises with larger n, larger α, or a true value farther from p_0 or μ_0.

6 CHOOSING THE PROCEDURE

Situation	Procedure
One group, yes/no data	one-proportion z
One group, numerical data	one-sample t
Two independent groups, yes/no	two-proportion z (CI or test)
Two independent groups, numerical	two-sample t
Same subjects measured twice / pairs	paired t on the differences
Slope of a regression line	$b \pm t^* SE_b$, $df = n - 2$