Confidence intervals

"We are 95% confident the true population proportion is in the range (0.4, 0.7)."

$$\hat{p} \pm z^* SE(\hat{p}) \qquad \bar{y} \pm t_{df}^* SE(\bar{y})$$

- Sample statistic
  - the actual observed value of the proportion ($\hat{p}$) or the mean ($\bar{y}$)
  - We use this as the best guess of the population parameter and add some give or take (aka Margin of Error)
- Critical value
  - A value looked up based on how confident we want to be
    - z is looked up based on normal distribution; t_{df} is looked up based on t-distribution and df
    - df is "degrees of freedom", here df = n - 1
  - Bigger confidence -> bigger critical value -> bigger margin of error -> wider CI
- Standard error
  - More evidence (aka large n) -> lower SE -> smaller margin of error
  - More variation -> higher SE -> larger margin of error

Confidence intervals

"We are 95% confident the true population mean is in the range (4, 10)."

$$\hat{p} \pm \underbrace{z^* SE(\hat{p})} \qquad \bar{y} \pm \underbrace{t_{df}^* SE(\bar{y})}$$

"Margin of Error"

- Affected by level of confidence (bigger confidence, bigger interval, 🧑‍🍳🍕)
- Affected by sample size (more evidence -> smaller ME -> smaller interval)

- CIs are the sample stat plus or minus ME, this means that the width of the CI is 2*ME (start at middle and take 1 ME sized step in both directions)
  - Given CI of (5, 15) we can find ME and sample stat with:
    - **2*ME = CI_{hi} - CI_{low}** -> 2*ME = 15 - 5 -> 2*ME = 10 -> ME = 5
    - **sample stat = CI_{hi} - ME** -> sample stat = 15 - 5 = 10

CI for population proportion

$$\hat{p} \pm \underbrace{z^* SE(\hat{p})}_{\text{Margin of Error}}$$

CI for population mean

$$\bar{y} \pm \underbrace{t_{df}^* SE(\bar{y})}_{\text{Margin of Error}}$$

Assumptions

- Success/failure
  - at least 10 successes ($\hat{p}$) and 10 failures ($\hat{q}$ -> aka 1 - $\hat{p}$)

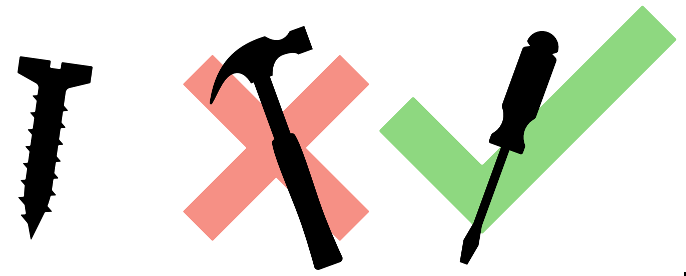
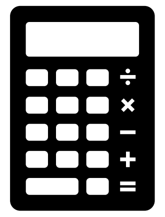
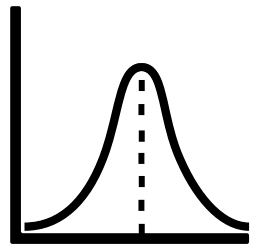
Assumptions

- Nearly normal
  - Distribution should resemble the normal distribution. When sample size is larger (40ish) we can have larger skew without worry.

Shared Assumptions

- Randomization
  - Sample is a random sample (we want to be representative, not biased)
- 10%
  - sample size (n) must be less than or equal to 10% of population

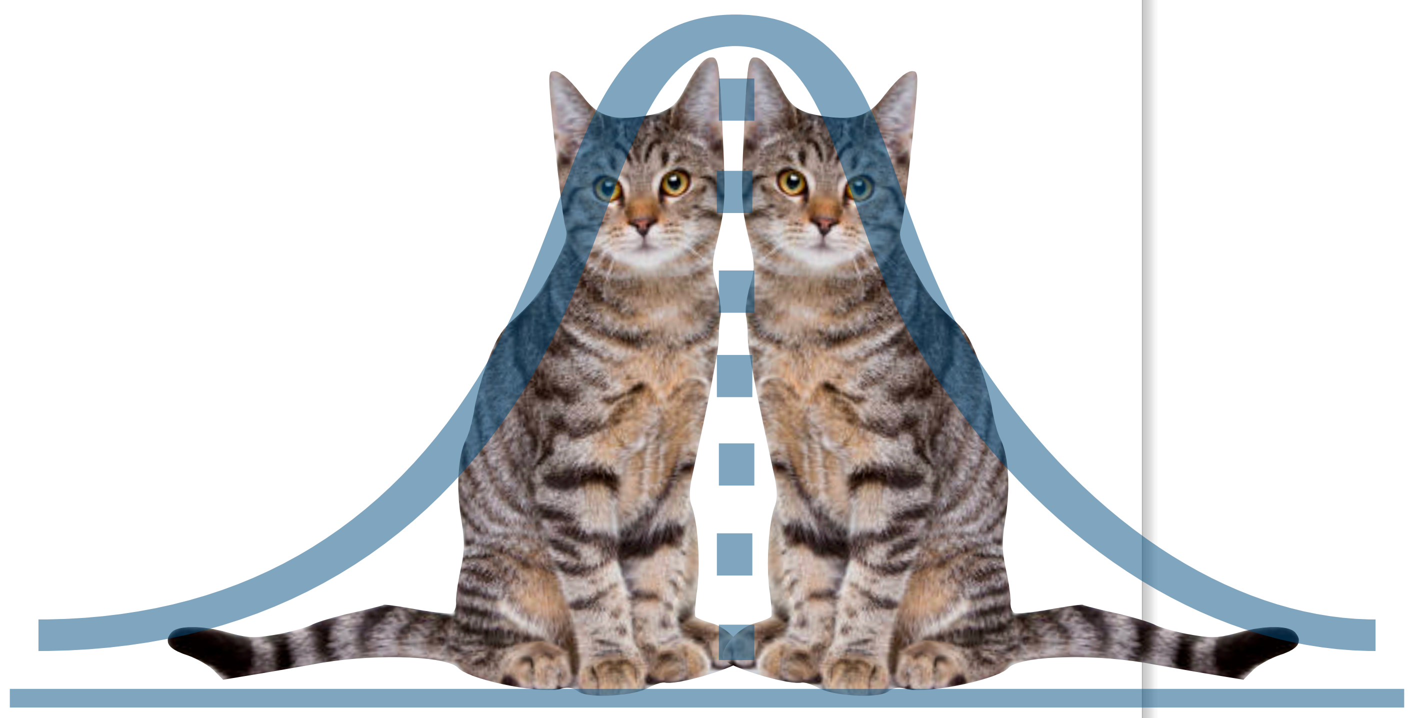
Hypothesis testing - steps

- **Hypotheses** H_0 & H_a
  - State the hypotheses (aka the 2 possible outcomes of the test)
- **Model** 
  - Check your assumptions (use the right tool for the job)
- **Mechanics** 
  - Formula time! (or JMP time)
- **Conclusions** 
  - Read the p-value - “If p is low, H_0 must go”
    - If p < alpha (aka “level of significance”) reject null hypothesis; there is evidence supporting H_a
    - If p > alpha (aka “level of significance”) fail to reject null hypothesis; there is not enough evidence supporting H_a

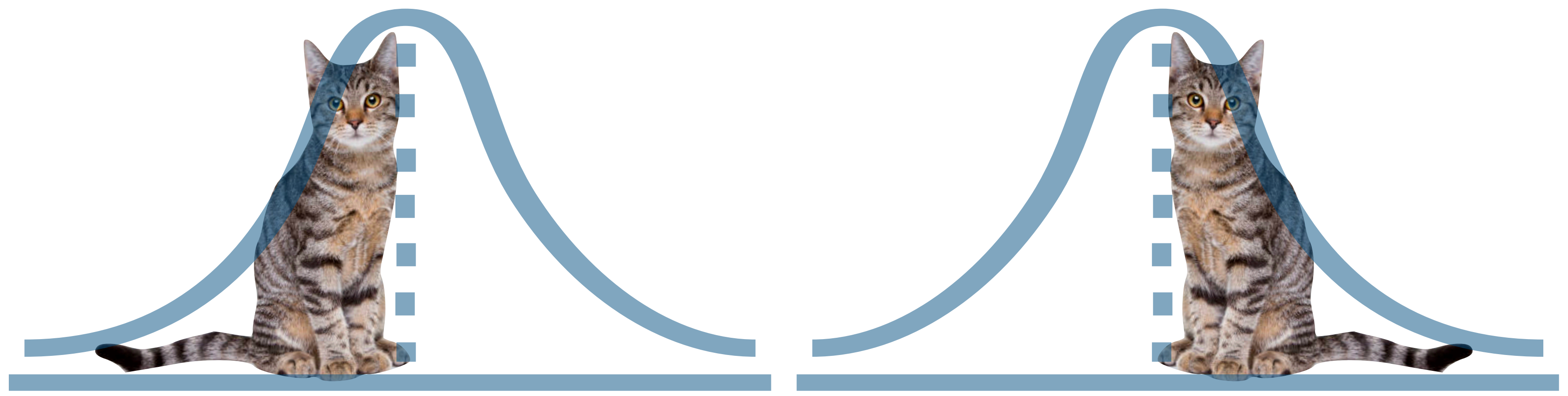
Hypothesis testing - hypotheses

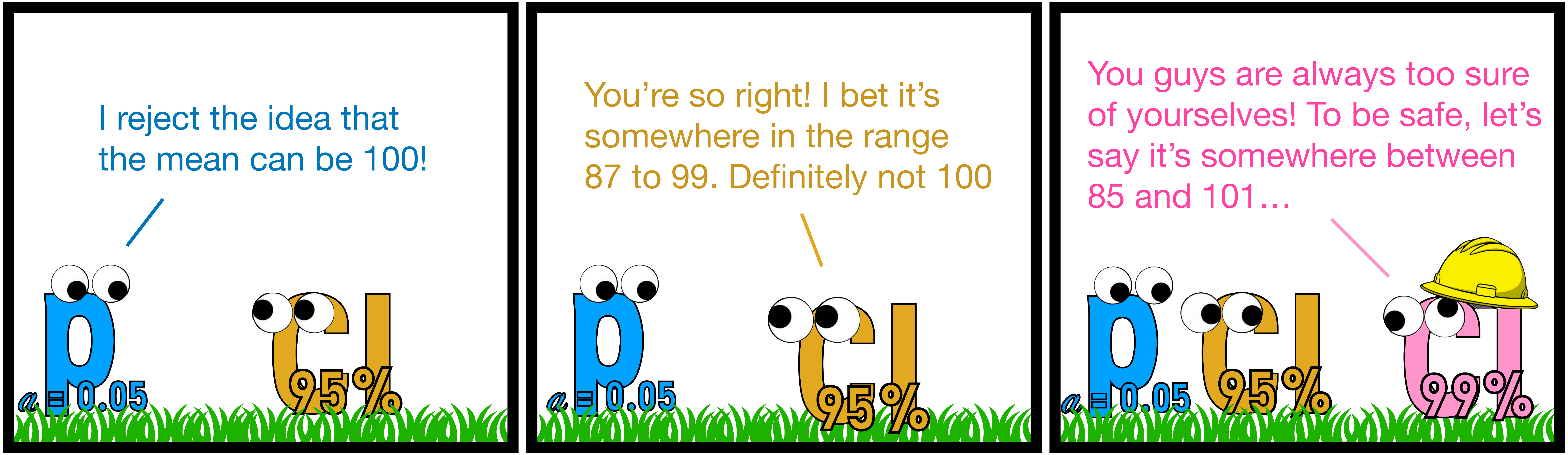
“We reject the hypothesis that the population mean is 0.”

- Example “null” hypotheses (H_0):
  - $\mu = 10$ “the true population mean is 10”
  - $p = 0.4$ “the true population proportion is 0.4”
  - $\mu_a - \mu_b = 0$ “the true difference in population means is 0”

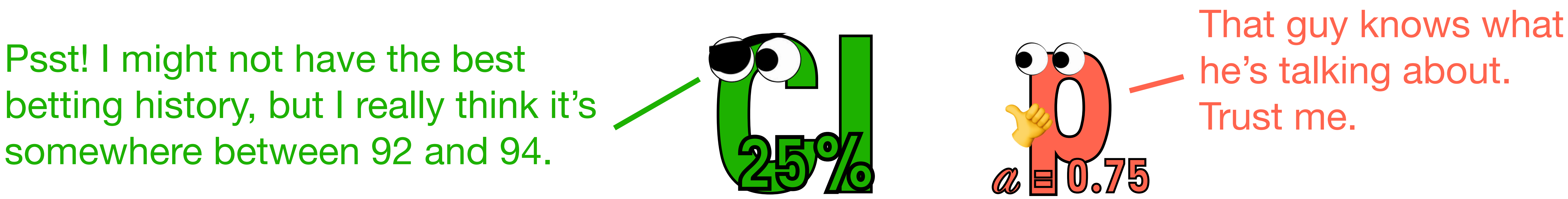


- Example “alternative” hypotheses (H_a):
  - “Two-tailed” tests - use both “tails” of distribution to measure probability
    - $\mu \neq 10; \quad p \neq 0.4; \quad \mu_a - \mu_b \neq 0$
  - “One-tailed” tests - use only 1 “tail” of distribution to measure probability
    - $\mu > 10; \quad p > 0.4; \quad \mu_a - \mu_b > 0$
    - $\mu < 10; \quad p < 0.4; \quad \mu_a - \mu_b < 0$

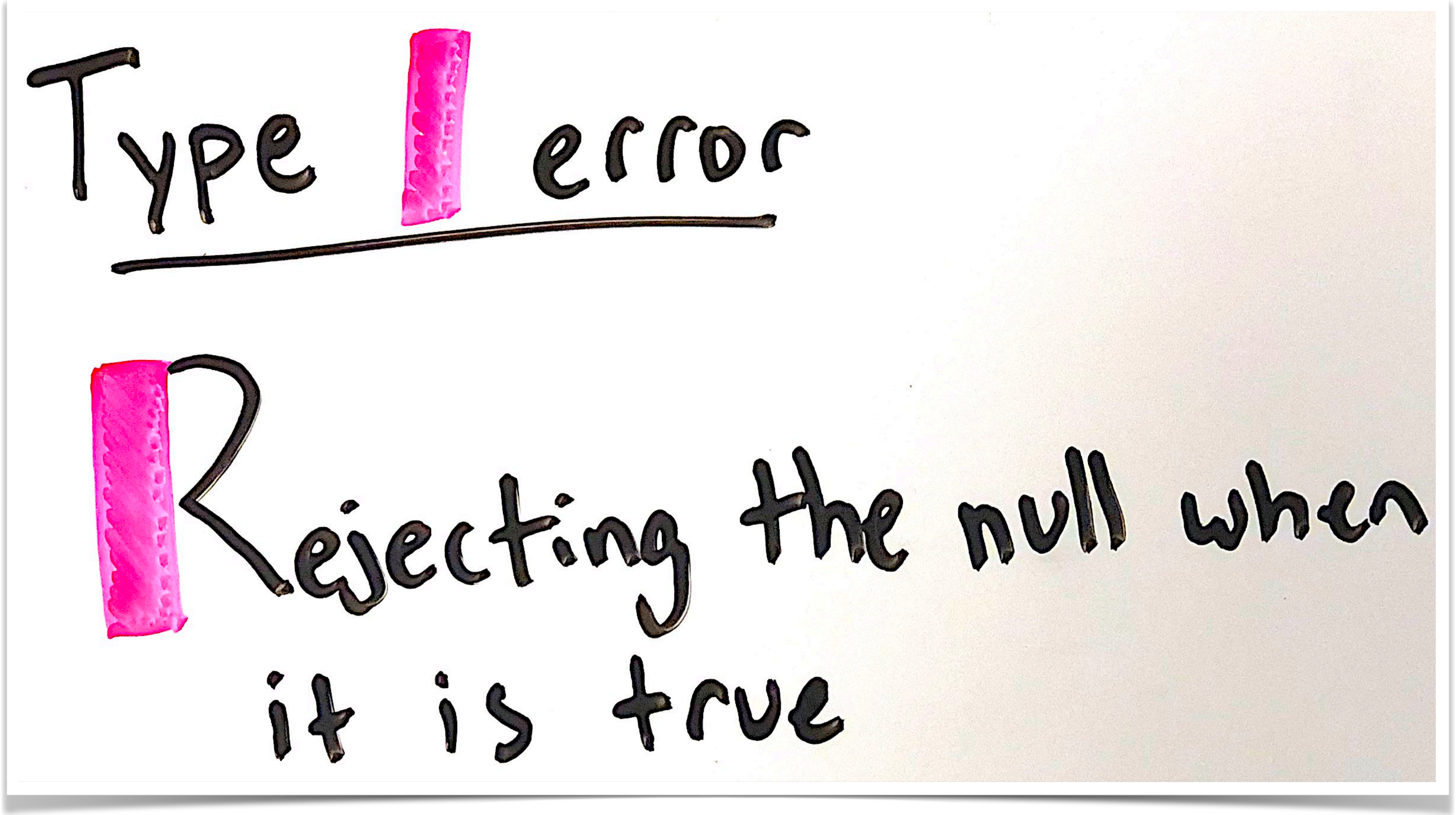


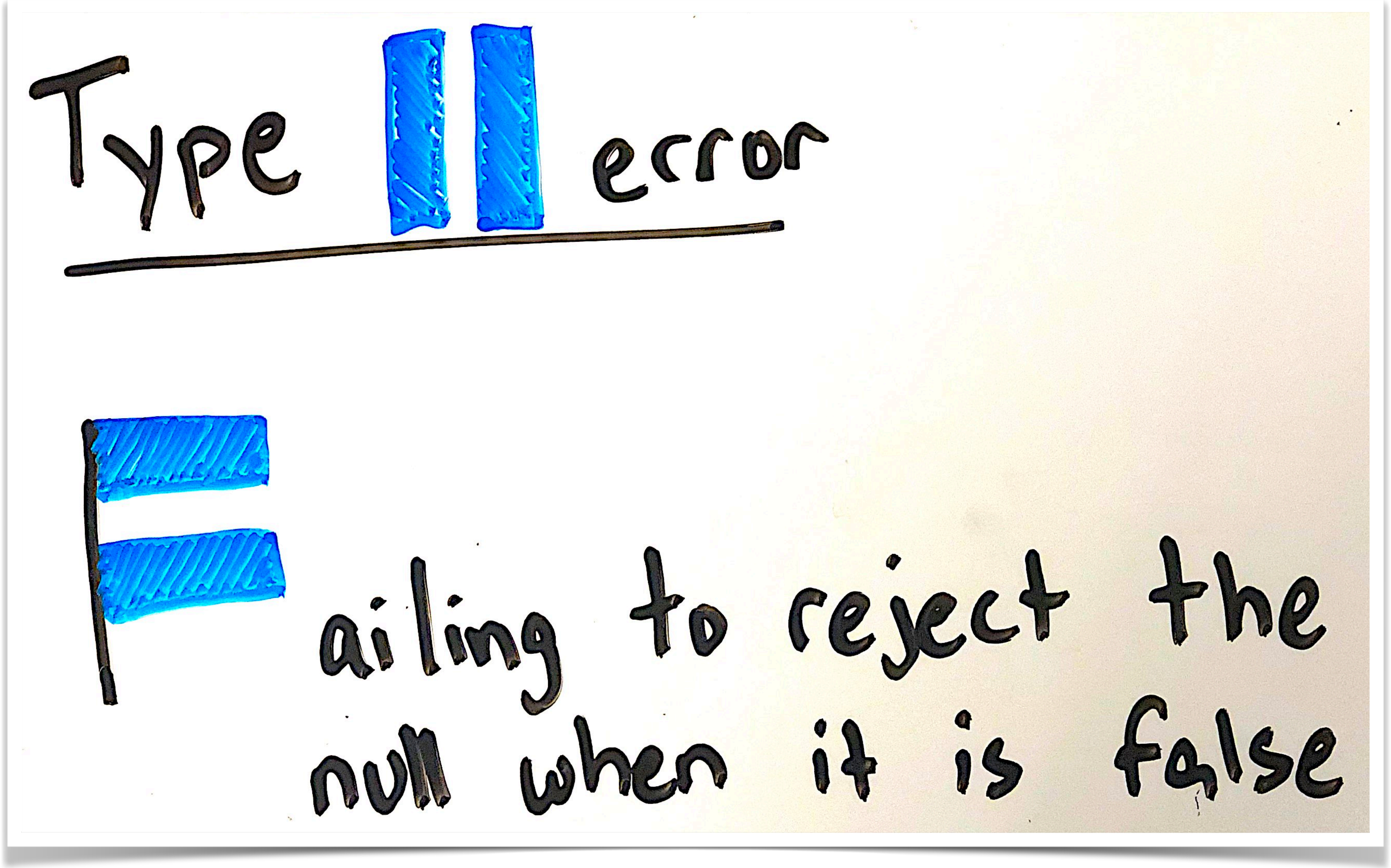


- **Tests and Confidence Intervals will agree*** 🤝
  - *requires alpha and confidence to match (eg *alpha = 0.05 & 95% CI* or *alpha = 0.01 and 99% CI*)
- **If the test rejects a possible value -> CI won't have that value in it's range**
  - If we reject null that mean is 10 then the CI's range will not include 10, maybe (5, 8)
  - If we fail to reject that the mean is 10 then the CI's range will include 10, maybe (1, 12)

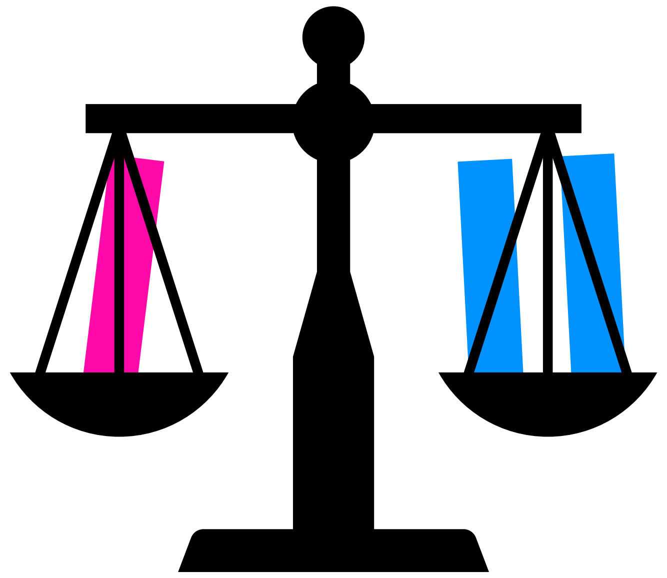


Hypothesis testing - errors





- **The probability of making a Type I is denoted as α** (aka alpha or “level of significance”)
  - When $p < \alpha$ we’re saying “the probability of this happening is very low if the null hypothesis is indeed true”, but rare events can just happen 🤷‍♂️ and this can lead to type I errors
- **The probability of making Type II is denoted as β** (aka beta)
- Managing the errors is a balancing act
  - If you lower P(Type I) you raise P(Type II)
  - If you raise P(Type I) you lower P(Type II)

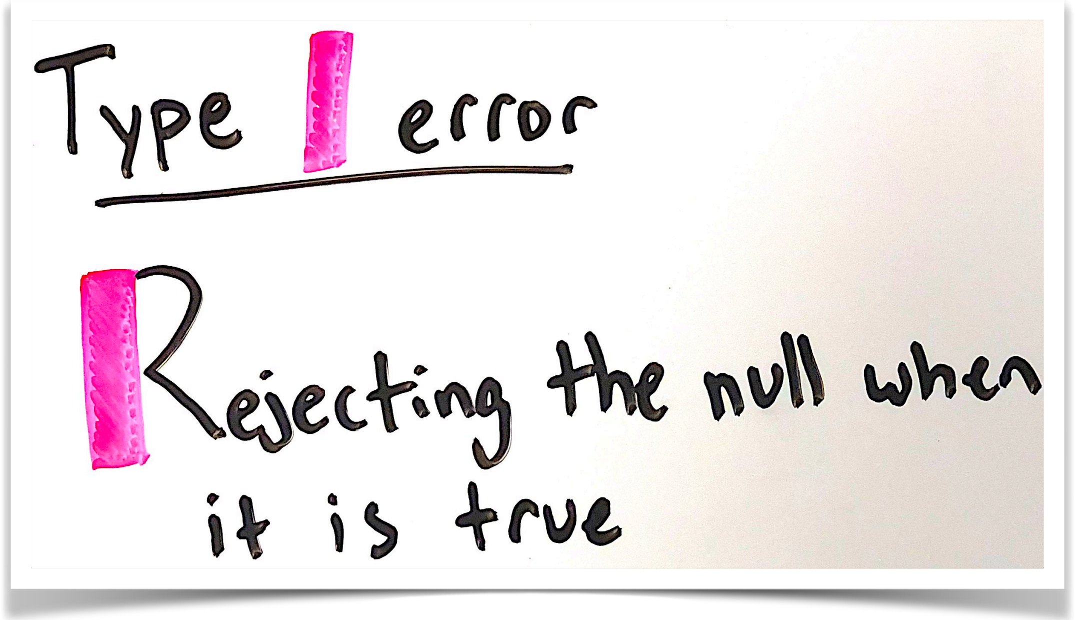


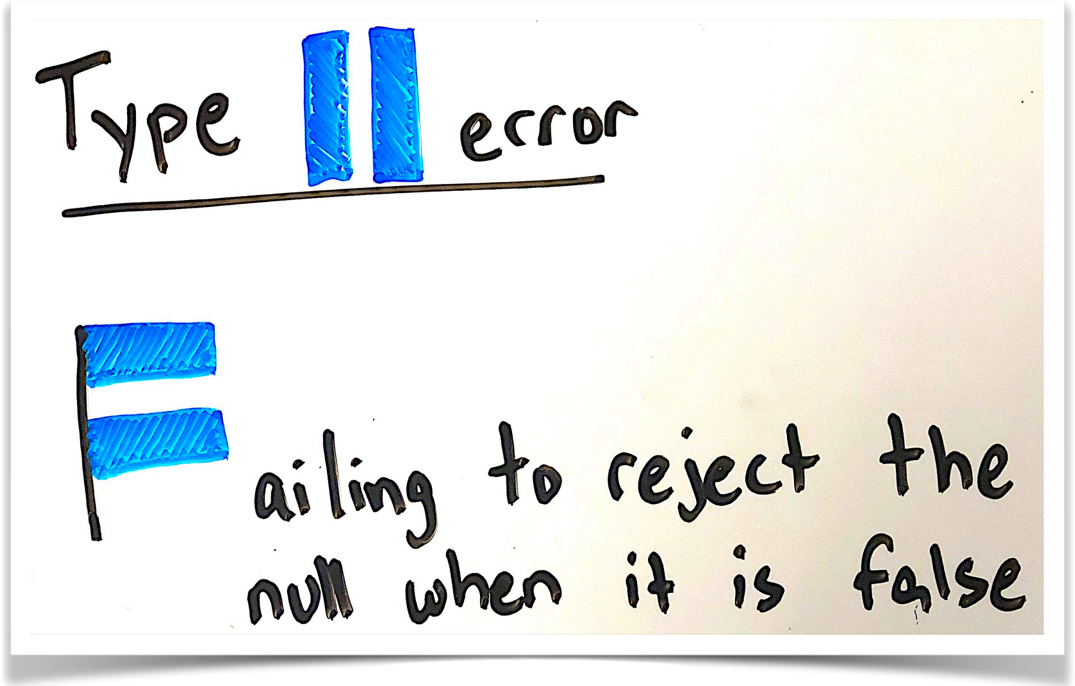
Hypothesis test utility belt

If p is low H_0 must go!

test	data	example null (fail to reject null if p-value > alpha)	example alts (reject null if p-value < alpha)	p-value comes from
Proportion z-test	1 categorical var w/2 categories	$p = 0.5$	$p \neq 0.5$ $p > 0.5$ $p < 0.5$	normal dist (z)
1-sample t-test	1 numeric var	$\mu = 10$	$\mu \neq 10$ $\mu > 10$ $\mu < 10$	student's t-dist (t_{df})
2-sample t-test	1 numeric var & 1 categorical var w/2 categories	$\mu_{group1} - \mu_{group2} = 0$	$\mu_{group1} - \mu_{group2} \neq 0$ $\mu_{group1} - \mu_{group2} > 0$ $\mu_{group1} - \mu_{group2} < 0$	student's t-dist (t_{df})
Chi-square test	2 categorical vars	Counts are independent	Counts are not independent	Chi-Square dist (χ^2_{df})

1. **Hypotheses**
2. **Model**
3. **Mechanics**
4. **Conclusions**

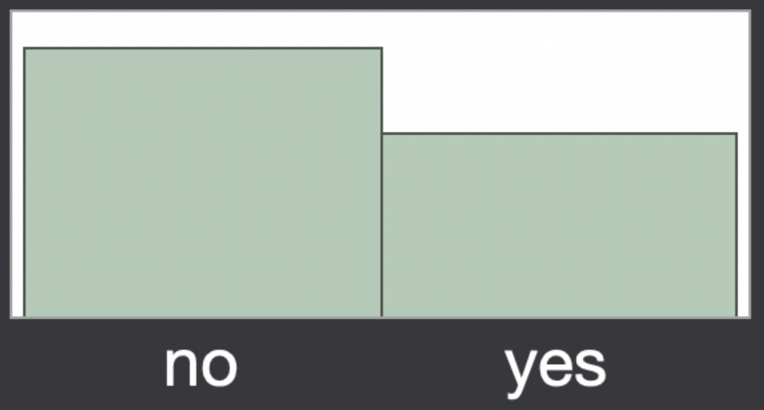




Make sure to review the assumptions of each!

- For all tests: Random & 10%
- For prop z test: Success/failure
- For t-tests: nearly normal (both groups for 2 sample) & independence (for 2 sample)
- For chi-square: expected cell frequency

is_automatic



Frequencies

Level	Count	Prob
no	19	0.59375
yes	13	0.40625
Total	32	1.00000

N Missing 0

2 Levels

Confidence Intervals

Level	Count	Prob	Lower CI	Upper CI	1-Alpha
no	19	0.59375	0.4226	0.744804	0.950
yes	13	0.40625	0.255196	0.5774	0.950
Total	32				

Note: Computed using score confidence intervals.

"We are 95% confident the true population proportion of manual cars is captured in the range (0.42, 0.74)."

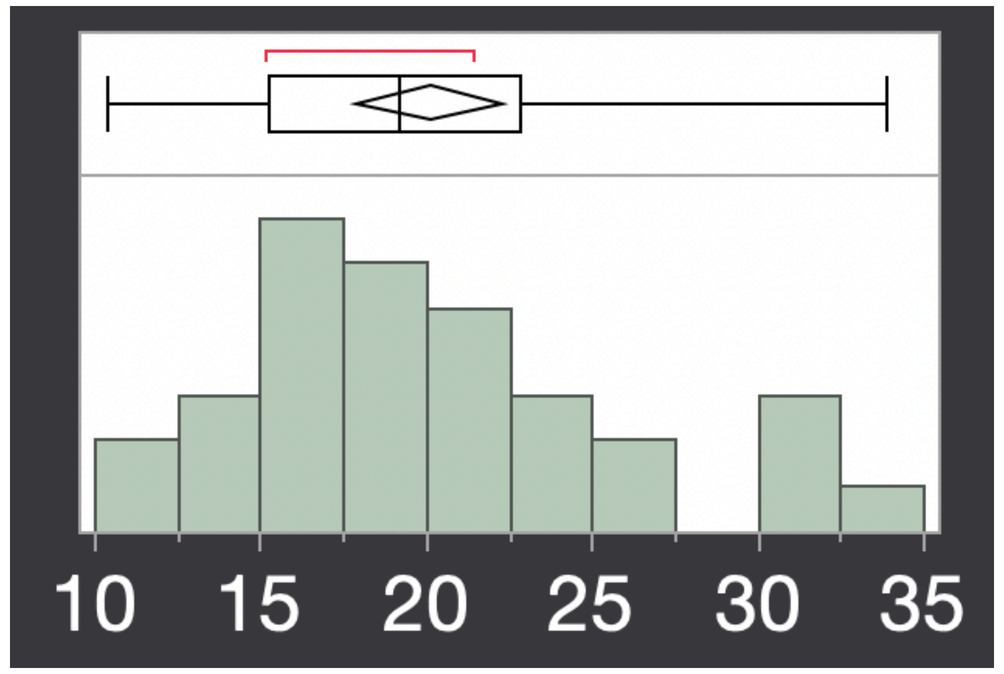
Interpret this one yourself for practice

95% confidence intervals. They will agree with a hypothesis test using alpha = 0.05

CI interpretation ✔-list

- Level of confidence (95%)
- Mention what the interval is for (population proportion)
- Context of data (manual cars)
- The range of the interval ((0.42, 0.74))

You don't have to say "captured in the range," but it's a pretty safe move



Quantiles

100.0%	maximum	33.9
99.5%		33.9
97.5%		33.9
90.0%		30.4
75.0%	quartile	22.8
50.0%	median	19.2
25.0%	quartile	15.275
10.0%		13.6
2.5%		10.4
0.5%		10.4
0.0%	minimum	10.4

Summary Statistics

Mean	20.090625
Std Dev	6.0269481
Std Err Mean	1.065424
Upper 95% Mean	22.263571
Lower 95% Mean	17.917679
N	32

"We are 95% confident the true population mean of mpg is captured in the range (17.9, 22.3)."

95% confidence intervals.
They will agree with a hypothesis test using alpha = 0.05

CI interpretation ✔-list

- Level of confidence (95%)
- Mention what the interval is for (population mean)
- Context of data (mpg)
- The range of the interval ((17.9, 22.3))

You don't have to say "captured in the range," but it's a pretty safe move

🚨 Verry important to look here! 🚨
Since we have `yes–no`:

- Positive numbers in CI indicate `yes>no` on average
- Dictates which p-value to use

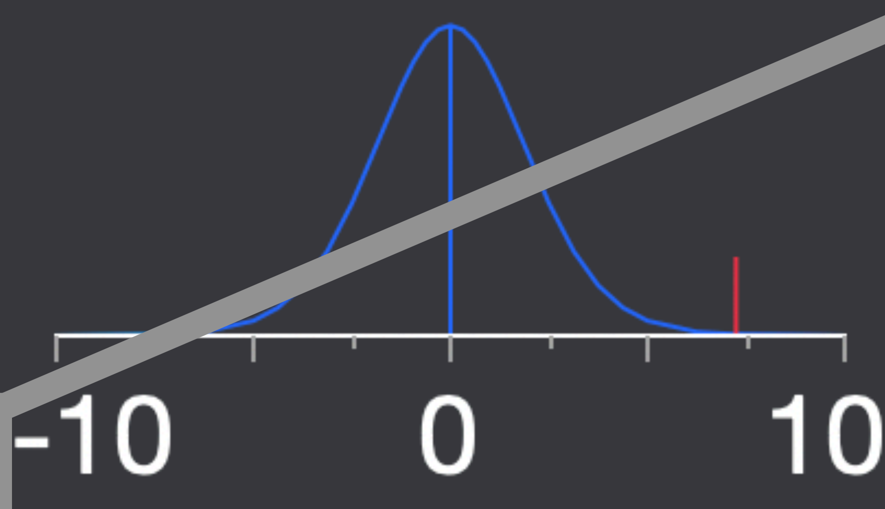
The observed difference in mean.
Based on the cars we had in the sample:
AVG(mpg_{yes}) - AVG(mpg_{no}) = 7.2
We'll use the CI and the hypothesis test to see what we can expect from the whole population

t Test

yes-no

Assuming unequal variances

Difference	7.2449	t Ratio	3.767123
Std Err Dif	1.9232	DF	18.33225
Upper CL Dif	11.2802	Prob > |t|	0.0014*
Lower CL Dif	3.2097	Prob > t	0.0007*
Confidence	0.95	Prob < t	0.9993



Use when: $H_a : \mu_{yes} - \mu_{no} \neq 0$
Use when: $H_a : \mu_{yes} - \mu_{no} > 0$
Use when: $H_a : \mu_{yes} - \mu_{no} < 0$
Match the order of subtraction with `yes–no`

"We reject the null hypothesis that there is no difference of means, there is evidence that the average for yes is greater than the average for no at the 0.05 level of significance ($p < \alpha$)

"We are 95% confident the true difference of means between manuals and automatics is captured in the range (3.2, 11.3). It appears automatics have higher mpg on average"

p-values! (aka **p**robability values) If our null hypotheses was $H_o : \mu_{yes} - \mu_{no} = 0$ (aka "the means are the same")
read these as "the **p**robability we'd see a difference this big in our sample assuming there is no difference (aka assuming H_0)"