

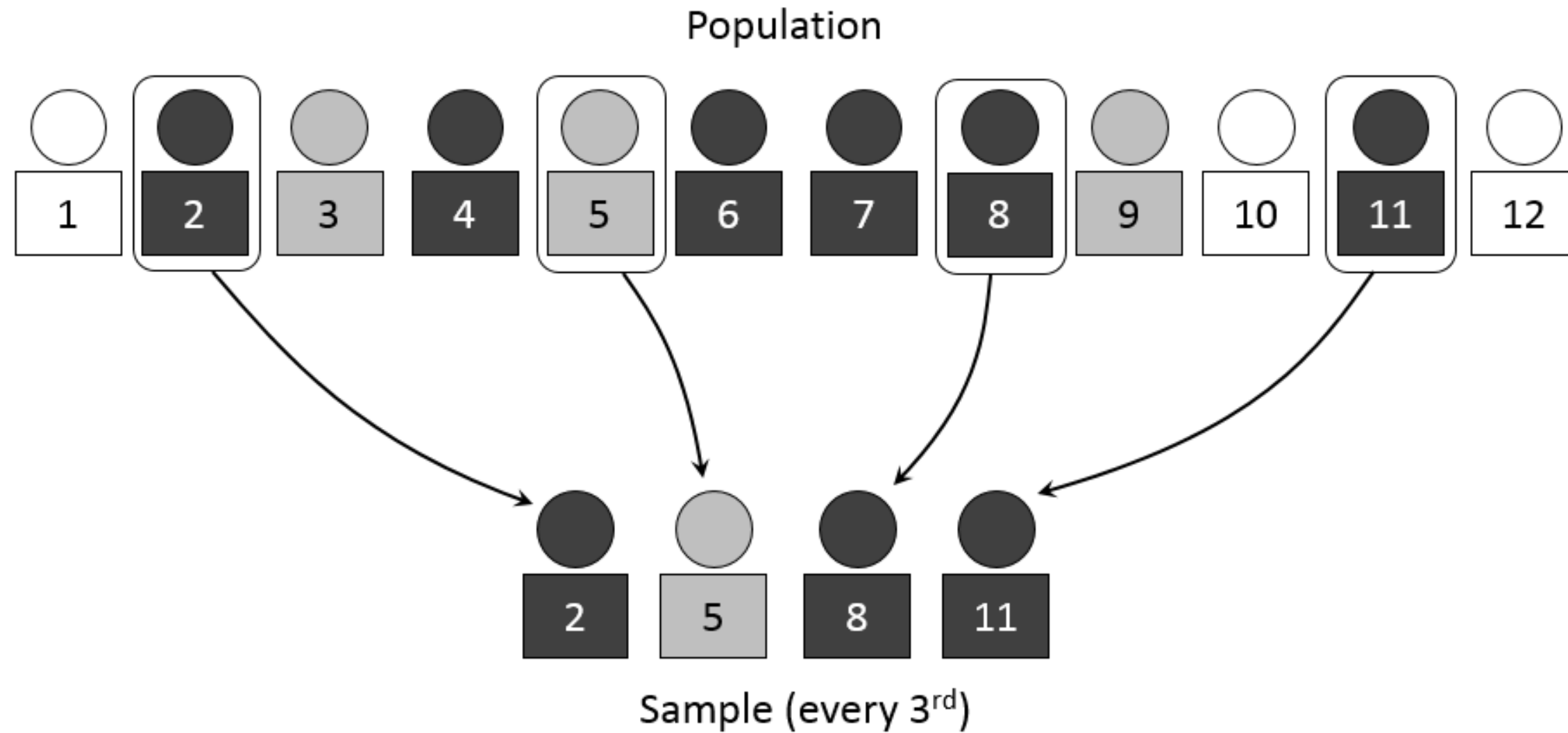
# Confidence intervals

PREPARING FOR STATISTICS INTERVIEW QUESTIONS IN PYTHON



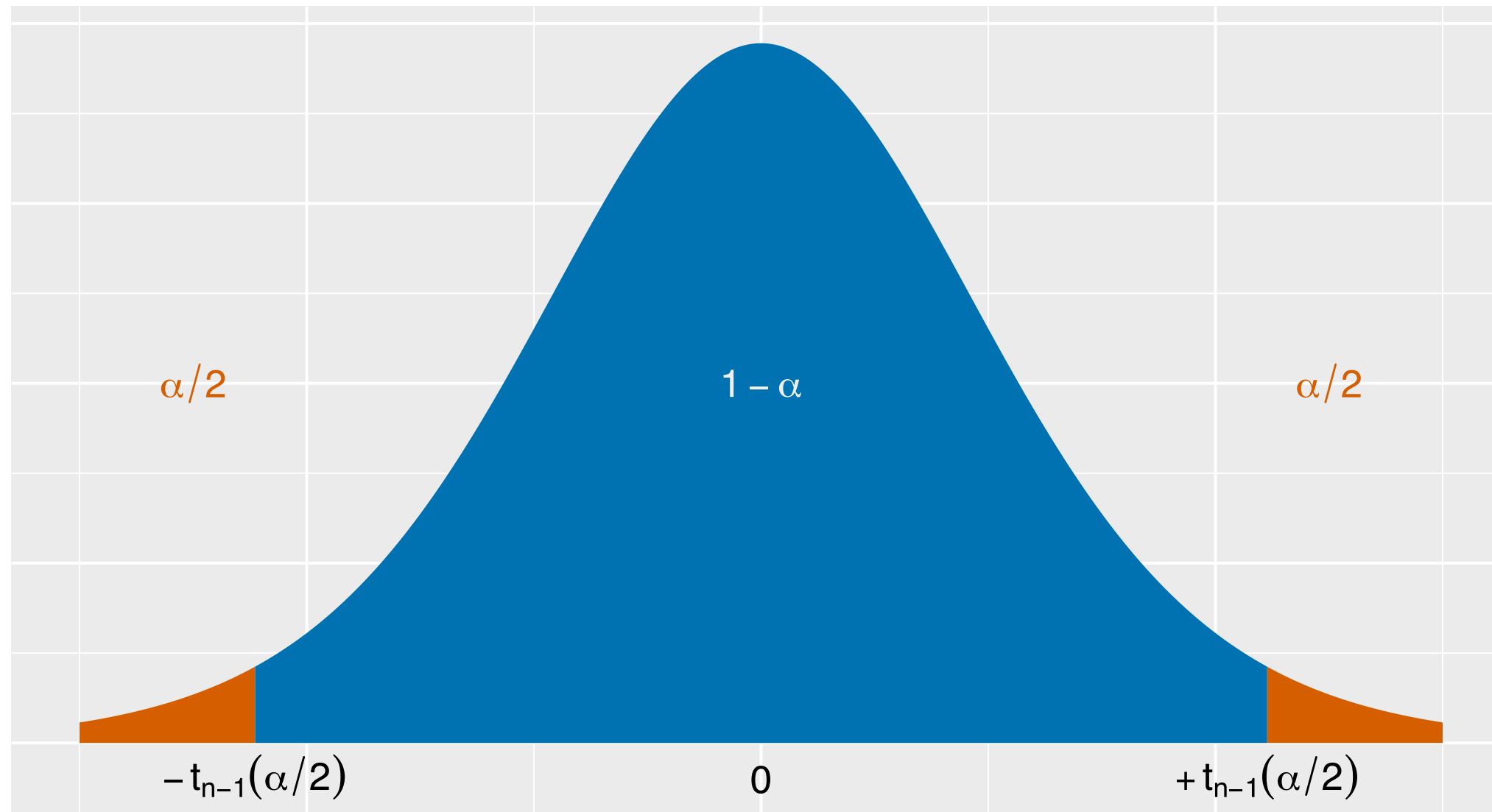
**Conor Dewey**  
Data Scientist, Squarespace

# Intro to sampling



<sup>1</sup> Wikimedia

# What is a confidence interval?



<sup>1</sup> Wikimedia

# Calculating confidence intervals

Means

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

# Calculating confidence intervals

Proportions

$$\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

# Example: means

```
import scipy.stats as st
a = range(10, 14)
st.t.interval(0.95, len(a) - 1, loc = np.mean(a),
              scale = st.sem(a))
```

```
(9.446, 13.554)
```

# Example: proportions

```
from sm.stats.proportion import proportion_confint  
proportion_confint(4, 10, .05)
```

```
(0.0964, 0.7036)
```

# Summary

- Sampling
- Confidence intervals
- Example



# Let's prepare for the interview!

PREPARING FOR STATISTICS INTERVIEW QUESTIONS IN PYTHON

# Hypothesis testing

PREPARING FOR STATISTICS INTERVIEW QUESTIONS IN PYTHON



**Conor Dewey**  
Data Scientist, Squarespace

# Quick review



<sup>1</sup> xkcd

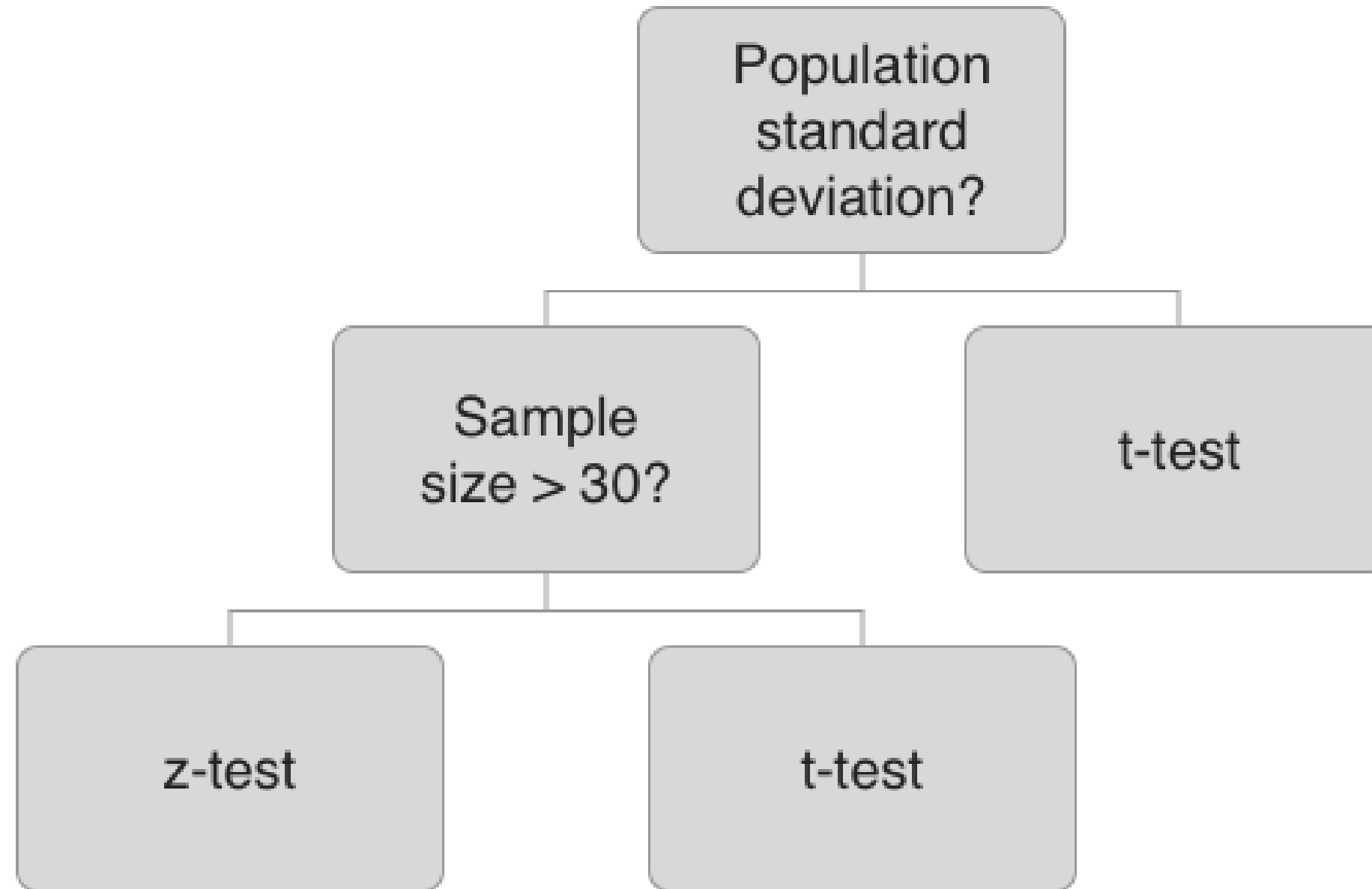
# Assumptions

- Random sampling
- Independent observations
- Normally distributed
- Constant variance

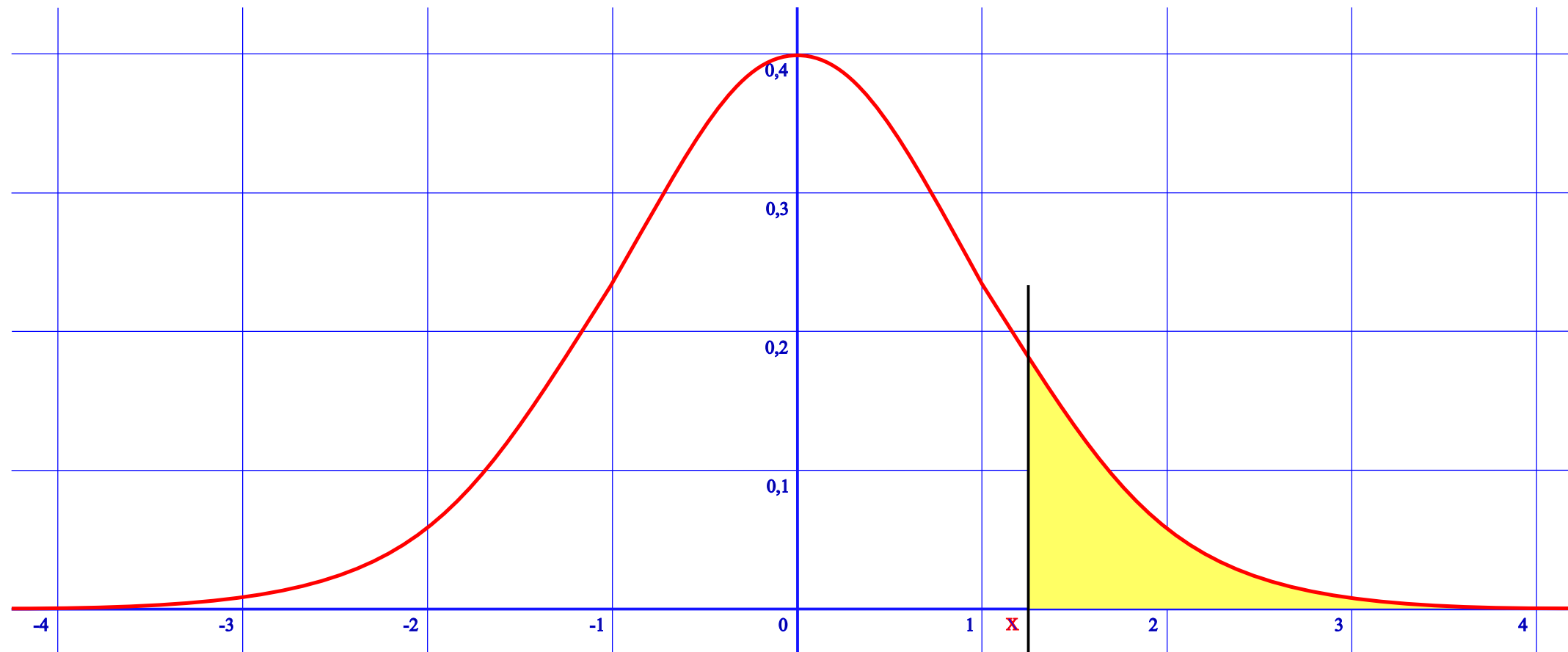
# Generating hypotheses

Two-tailed test	One tailed test
$H_0$ : Estimate = value	$H_0$ : Estimate $\geq$ value (Estimate $\leq$ value)
$H_1$ : Estimate $\neq$ value	$H_1$ : Estimate < value (Estimate > value)

# Which test to use



# Evaluating results



<sup>1</sup> Wikimedia

# Types of errors

		Reality	
		True	False
Measured or Perceived	True	Correct 😊	<b>Type 1 error</b> False Positive
	False	<b>Type 2 error</b> False Negative	Correct 😊

<sup>1</sup> AB Tasty



# Summary

- Quick review
- Assumptions
- Testing process
- Types of errors

# Let's prepare for the interview!

PREPARING FOR STATISTICS INTERVIEW QUESTIONS IN PYTHON

# Power and sample size

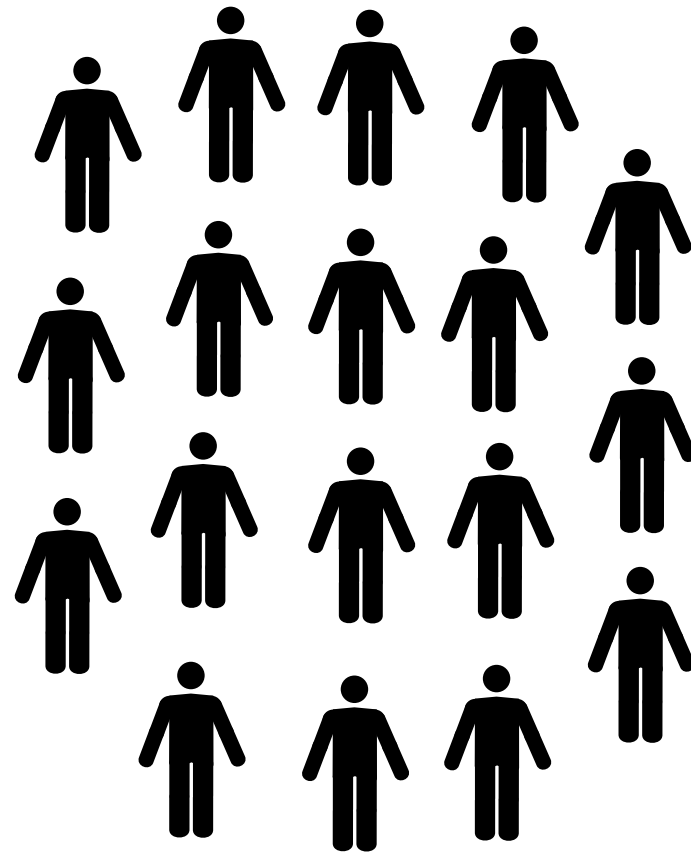
PREPARING FOR STATISTICS INTERVIEW QUESTIONS IN PYTHON



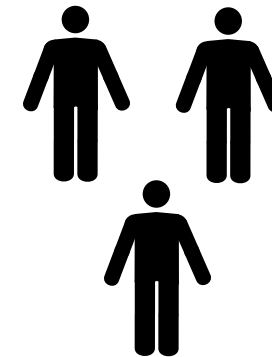
**Conor Dewey**  
Data Scientist, Squarespace

# Power analysis

Many



Few



<sup>1</sup> Public domain vectors

# Moving parts

- Effect size
- Significance level
- Power
- Sample size

# Calculating sample size

- `zt_ind_solve_power()`
- `tt_ind_solve_power()`
- `proportion_effectsize()`

# Example: conversion rates

```
from statsmodels.stats.power import zt_ind_solve_power
import statsmodels.stats.proportion as prop

std_effect = prop.proportion_effectsize(.20, .25)
zt_ind_solve_power(effect_size=std_effect, nobs1=None,
                    alpha=.05, power=.80)
```

```
1091.8962
```

# Example: conversion rates

```
from statsmodels.stats.power import zt_ind_solve_power
import statsmodels.stats.proportion as prop

std_effect = prop.proportion_effectsize(.20, .25)
zt_ind_solve_power(effect_size=std_effect, nobs1=None,
                    alpha=.05, power=.95)
```

```
1807.76215
```



# Summary

- Power analysis
- Moving parts
- Example

# Let's prepare for the interview!

PREPARING FOR STATISTICS INTERVIEW QUESTIONS IN PYTHON

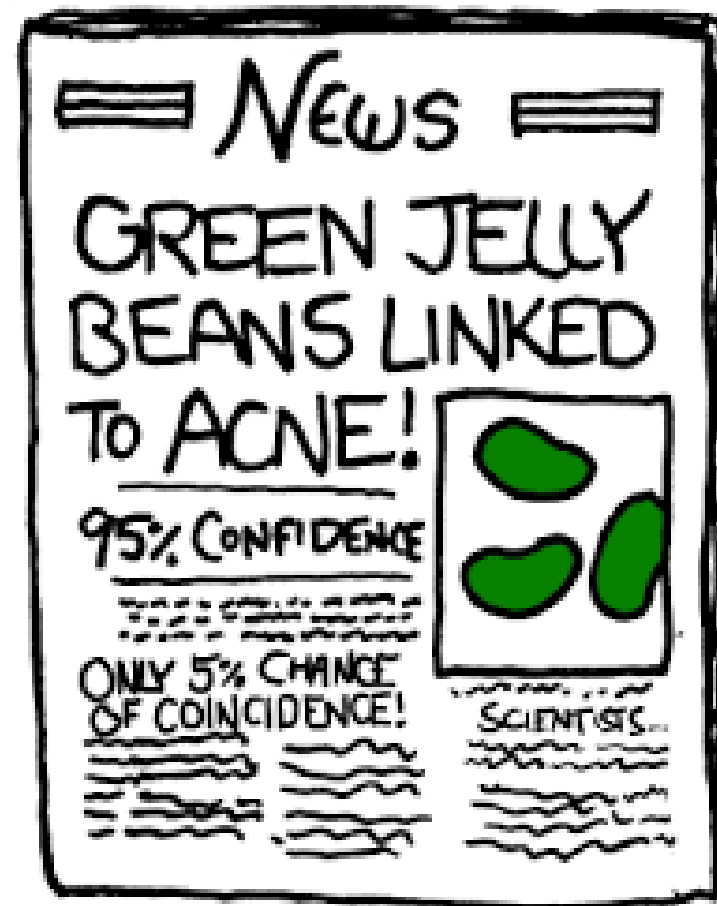
# Multiple testing

PREPARING FOR STATISTICS INTERVIEW QUESTIONS IN PYTHON



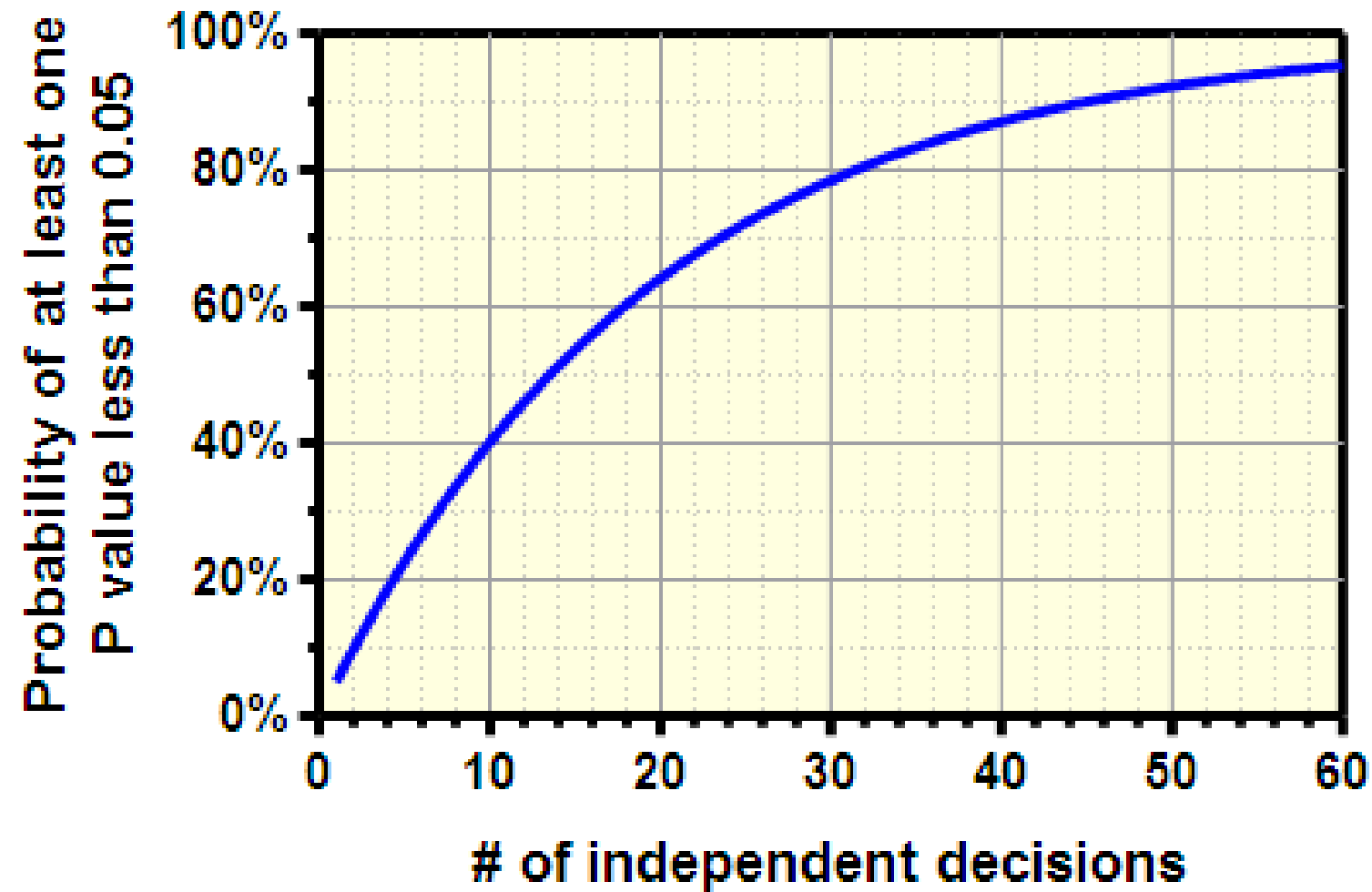
**Conor Dewey**  
Data Scientist, Squarespace

# Multiple comparisons problem



<sup>1</sup> xkcd

# Correcting for multiple comparisons



<sup>1</sup> GraphPad

# Common approaches

- Bonferroni correction
- Sidak correction
- Step-based procedures
- Tukey's procedure
- Dunnet's correction

# Bonferroni correction

*The original p value* 

$$\text{Bonferroni-corrected } p \text{ value} = \frac{\alpha}{n}$$

*The number of tests performed* 

<sup>1</sup> Wikimedia

# Example

```
from statsmodels.sandbox.stats.multicomp import multipletests
p_adjusted = multipletests(pvals, alpha=.05, method='bonferroni')
print(p_adjusted[0])
print(p_adjusted[1])
```

```
[ True False False False False]
[0.05 0.25 0.5  1.   1.   ]
```



# Side effects



<sup>1</sup> What's wrong with Bonferroni adjustments

# Summary

- Multiple comparisons problem
- Common correction approaches
- Bonferroni correction

# Let's prepare for the interview!

PREPARING FOR STATISTICS INTERVIEW QUESTIONS IN PYTHON