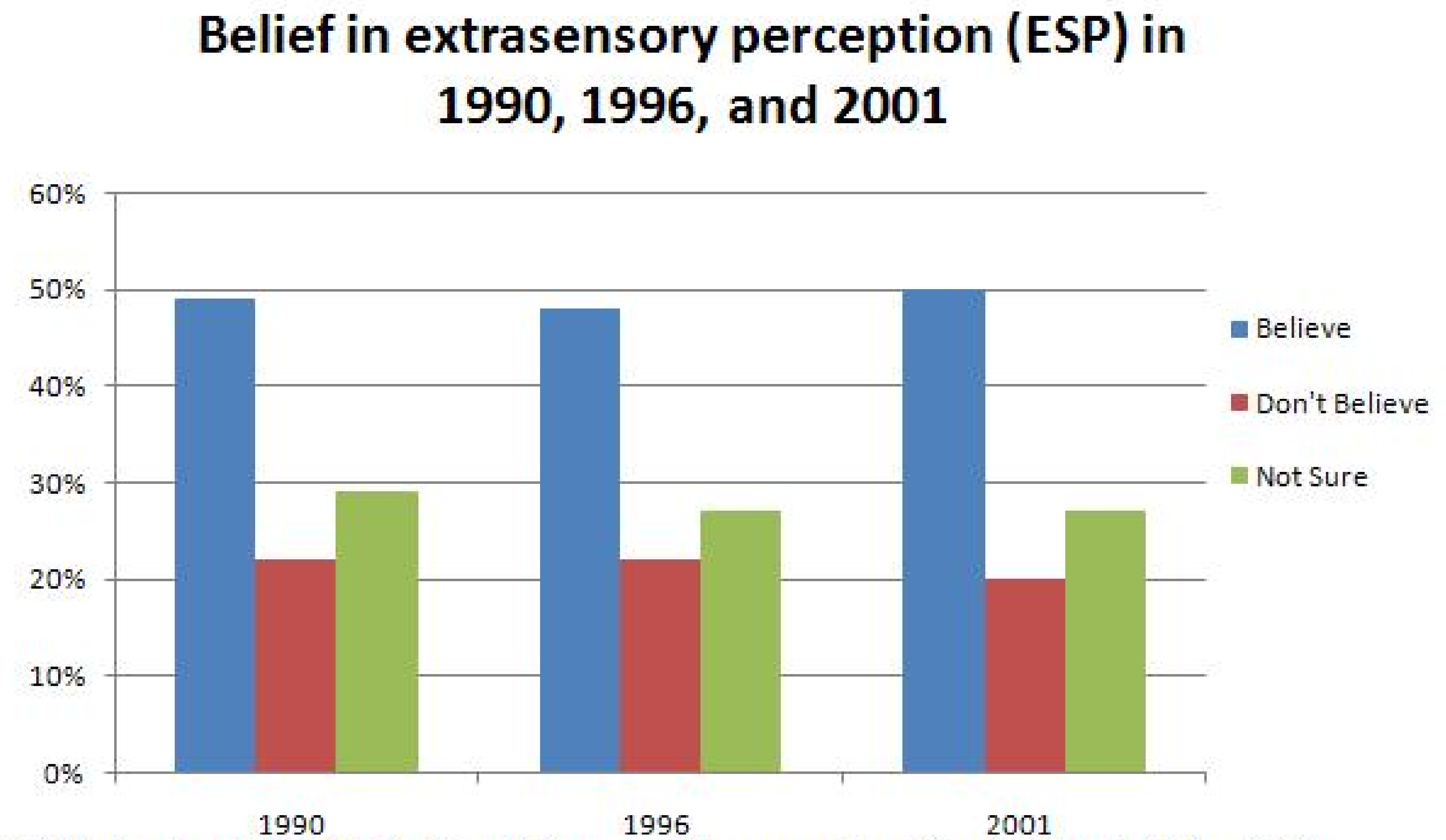
Statistical Hypothesis Testing

Stat 120

April 18 2022

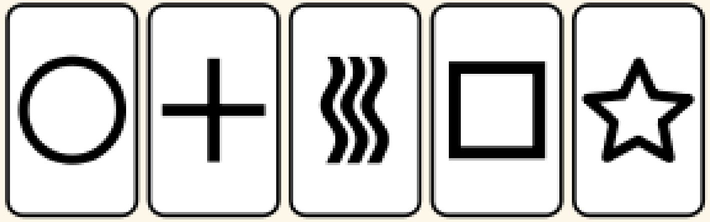
Extrasensory Perception



SOURCE: Americans' Belief in Psychic and Paranormal Phenomena Is up Over Last Decade (8 June 2001), http://www.gallup.com/poll/4483/Americans-Belief-Psychic-Paranormal-Phenomena-Over-Last-Decade.aspx

Extrasensory Perception

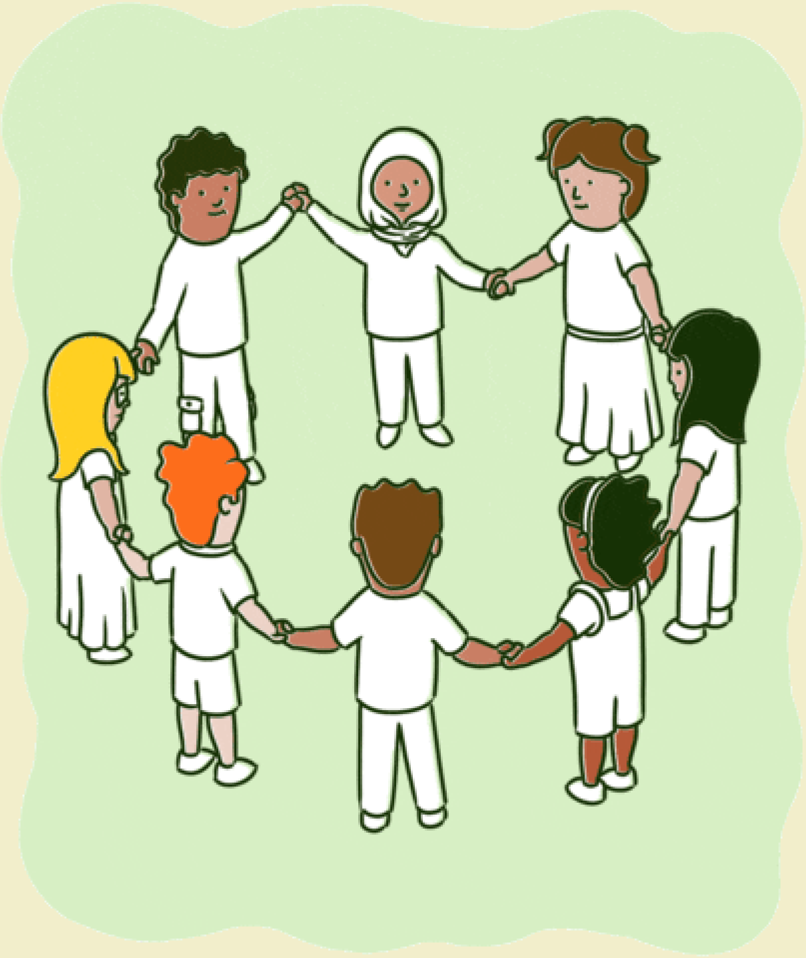
One way to test for ESP is with Zener cards:



Subjects draw a card at random and telepathically communicate this to someone who then guesses the symbol

Your Turn 1

05:00



Randomly choose a letter from A B C D E and write it down (don't show anyone!)

Find a partner, telepathically communicate your letter (no auditory or visual clues!) and have them guess your letter.

Repeat a couple of times then switch roles.

How often did you guess correctly?

Suppose you did this 10 times and guessed correctly 3 times. Is this evidence that you have ESP abilities?

Extrasensory Perception

There are five cards with five different symbols. If there is no such thing as ESP, what proportion p of guesses should be correct?

1. $p = 0$
2. $p = 1/4$
3. $p = 1/5$
4. $p = 1/2$

▶ Click for answer

Extrasensory Perception (Example 1)

Let $\hat{p}$ denote the sample proportion of correct guesses. Which of the statistics below would give the strongest evidence for ESP?

1. $\hat{\mathrm{p}} = 0$
2. $\hat{\mathrm{p}} = 1/5$
3. $\hat{\mathrm{p}} = 1/2$
4. $\hat{\mathrm{p}} = 3/4$

▶ Click for answer

Extrasensory Perception

- As we've learned, statistics vary from sample to sample
- Even if the "population/true" proportion is $p = 1/5$, not every sample proportion will be exactly $1/5$

How do we determine when a sample proportion is far enough above 1/5 to provide evidence of ESP?

Statistical Test

A statistical test uses data from a sample to assess a claim about a population or experiment

Null Hypothesis: (H_0) Claim that there is no effect or difference.

Alternative Hypothesis: (H_a) Claim for which we seek evidence.

Always claims about **population parameters.**

ESP Hypothesis

For the ESP experiment:

- $H_0 : p = 1/5$
- $H_a : p > 1/5$

Helpful hints:

- H_0 usually includes $=$
- H_a usually includes $>$, $<$, or $\neq$
- The direction in H_a depends on the question being asked, not based on what the data shows!
- The data should be used as an evidence supporting or refuting H_a.

Sleep Vs. Caffeine (Example 2)

Students were given words to memorize, then randomly assigned to take either a 90 min nap, or a caffeine pill. $2\frac{1}{2}$ hours later, they were tested on their recall ability.

- Explanatory variable: **sleep or caffeine**
- Response variable: **number of words recalled**

Is sleep or caffeine better for memory?

Mednick, Cai, Kanady, and Drummond (2008). "Comparing the benefits of caffeine, naps and placebo on verbal, motor and perceptual memory," Behavioral Brain Research, 193, 79-86.

Sleep Vs. Caffeine

What is the parameter of interest in the sleep versus caffeine experiment?

1. Proportion
2. Difference in proportions
3. Mean
4. Difference in means
5. Correlation

▶ Click for answer

Sleep Vs. Caffeine

- Let μ_{s} and μ_{c} be the mean number of words recalled after sleeping and after caffeine.
- Is there a difference in average word recall between sleep and caffeine?

Sleep Vs. Caffeine

- What are the null and alternative hypothesis?

1. $\mathrm{H}_0 : \mu_\mathrm{s} \neq \mu_\mathrm{c}, \mathrm{H}_\mathrm{a} : \mu_\mathrm{s} = \mu_\mathrm{c}$
2. $\mathrm{H}_0 : \mu_\mathrm{s} = \mu_\mathrm{c}, \mathrm{H}_\mathrm{a} : \mu_\mathrm{s} \neq \mu_\mathrm{c}$
3. $\mathrm{H}_0 : \mu_\mathrm{s} \neq \mu_\mathrm{c}, \mathrm{H}_\mathrm{a} : \mu_\mathrm{s} > \mu_\mathrm{c}$
4. $\mathrm{H}_0 : \mu_\mathrm{s} = \mu_\mathrm{c}, \mathrm{H}_\mathrm{a} : \mu_\mathrm{s} > \mu_\mathrm{c}$
5. $\mathrm{H}_0 : \mu_\mathrm{s} = \mu_\mathrm{c}, \mathrm{H}_\mathrm{a} : \mu_\mathrm{s} < \mu_\mathrm{c}$

▶ Click for answer

Difference in Hypothesis

Note: the following two sets of hypotheses are equivalent, and can be used interchangeably:

$$\begin{aligned} \mathrm{H_0} &: \mu_1 = \mu_2 & \mathrm{H_0} &: \mu_1 - \mu_2 = 0 \\ \mathrm{H_a} &: \mu_1 \neq \mu_2 & \mathrm{H_a} &: \mu_1 - \mu_2 \neq 0 \end{aligned}$$

Guess the best inference method

Options: TEST, CONFIDENCE INTERVAL, NEITHER

(a) What proportion of US adults support gun control?

▸ Click for answer

Guess the best inference method

Options: TEST, CONFIDENCE INTERVAL, NEITHER

(b) Does the proportion of US adults who support gun control differ between males and females?

▶ Click for answer

Guess the best inference method

Options: TEST, CONFIDENCE INTERVAL, NEITHER

(c) What proportion of this class supports gun control?

- Click for answer

Guess the best inference method

Options: TEST, CONFIDENCE INTERVAL, NEITHER

(d) How much more do men earn, on average, compared to women in the US?

- Click for answer

Guess the best inference method

Options: TEST, CONFIDENCE INTERVAL, NEITHER

(e) What proportion of Minnesota voters in the 2012 election voted for President Biden?

- Click for answer

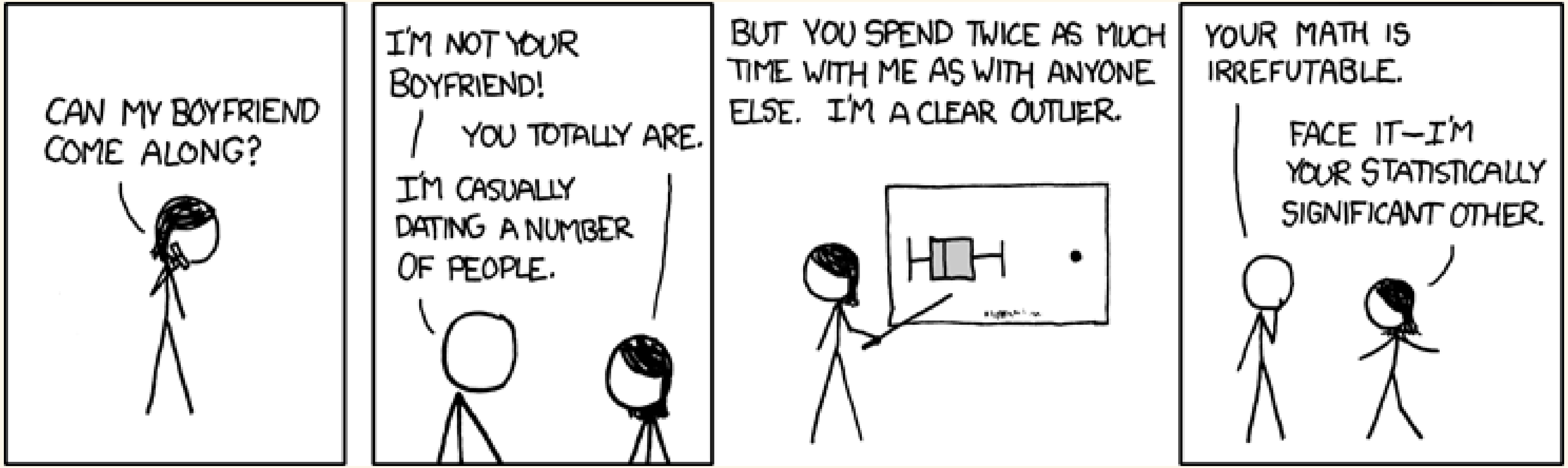
Guess the best inference method

Options: TEST, CONFIDENCE INTERVAL, NEITHER

(f) Is a higher rate of cricket chirping associated with higher summer night temps?

- Click for answer

Statistical Significance



xkcd.com

Statistical Significance

When results as extreme as the observed sample statistic are unlikely to occur by **random chance** alone (assuming the null hypothesis is true), we say the sample results are **statistically significant**

- If our sample is **statistically significant**, we have convincing **evidence against** H_0, **in favor of** H_a
- If our sample is **not statistically significant**, our test is **inconclusive**. The null hypothesis may be true (or maybe not).

Extrasensory Perception

$p =$ Proportion of correct guesses

$$\mathrm{H_0} : \mathrm{p} = 1/5$$
$$\mathrm{H_a} : \mathrm{p} > 1/5$$

If results are statistically significant ...

- the sample proportion of correct guesses is higher than is likely just by random chance (if ESP does not exist and $\mathrm{p} = 1/5$)
- we have evidence that the true proportion of correct guesses really is higher than $1/5$, and thus have evidence of ESP

Extrasensory Perception

$p =$ Proportion of correct guesses

$$\mathrm{H_0 : p = 1/5}$$
$$\mathrm{H_a : p > 1/5}$$

If results are **NOT** statistically significant ...

- the sample proportion of correct guesses could easily happen just by random chance (if ESP does not exist and $\mathrm{p = 1/5}$)
- we do not have enough evidence to conclude that $p > 1/5$, or that ESP exists
- BUT we still can't say that $p = 1/5$

Sleep Vs Caffeine

- μ_s and μ_c : mean number of words recalled after sleeping and after caffeine
- $H_0 : \mu_s = \mu_c$ and $H_a : \mu_s \neq \mu_c$ sleeping and after caffeine
- The sample difference in means is $\overline{x}_S - \overline{x}_C = 3$, and this is statistically significant.

Sleep Vs Caffeine

The sample difference in means is $\overline{x}_S - \overline{x}_C = 3$, and this is statistically significant.

We can conclude ...

1. there is a difference between sleep and caffeine for memory (and data show sleep is better)
2. there is a difference between sleep and caffeine for memory (and data show caffeine is better)
3. there is not a difference between sleep and caffeine for memory
4. nothing

Statistical Significance

- Hypothesis testing is similar to how our justice system works (or is suppose to work).

$$\mathrm{H}_0 : \text{Defendant is innocent vs. } \mathrm{H}_a\text{: Defendant is guilty}$$

Assumption: Defendant is innocent (H_0)

Verdict:

- **Guilty:** evidence (data) “beyond a reasonable doubt” points to guilt (Statistically significant)
- **Not Guilty:** evidence (data) not beyond a reasonable doubt, but we don't know if they are truly innocent (H_0)

BUT..

How do we determine statistical significance??

For ESP example:

- If there is no ESP, how unusual would it be to get 3 correct guesses in 10 tries?

For Sleep versus Caffeine example:

- If the effect of sleep and caffeine on recall is the same, how rare would it be to get an average difference of 3 words in the experiment conducted?

We assess this with a probability that we call a "p-value."

Summary

- Statistical tests use data from a sample to assess a claim about a population
- Statistical tests are usually formalized with competing hypotheses:
- Null hypothesis (H_0) : no effect or no difference
- Alternative hypothesis (H_a) : what we seek evidence for
- If data are statistically significant, we have convincing evidence against the null hypothesis, and in favor of the alternative