Understanding NHST and *p*-values

From Cassidy et al. (2019). Failing grade: 89% of Introduction-to-Psychology textbooks that define or explain statistical significance do so incorrectly. Advances in Methods and Practices in Psychological Science, Vol. 2(3), pp. 233-239.

Populations vs Samples

- Most researchers are interested in what is true at the population level rather than at their sample's level. Findings are typically only interesting if they generalize beyond the participants in a sample. A variety of statistical tools help researchers make inferences about the population based on their sample's data.

- Some researchers are interested in assessing whether the population effect might be a value other than exactly zero.

- These researchers often use Null Hypothesis Significance Testing (NHST) with *p*-values as a tool to help them with this assessment. Though as we see later this has become a controversial tool.

- The next few slides illustrate **conceptually** the logic behind NHST in the context of a correlation without getting into the detailed math (i.e., *t*-distributions, etc.). Our example is one in which we are not sure if there is a positive or negative correlation between our variables (i.e., a two-sided test based on a non-directional hypothesis).

Scenario

Consider a situation in which you are a researcher interested in the correlation between weight (lbs) and height (inches) for a population of 100,000 people.

Unfortunately, you only have a random sample of 10 people from the population to assess this relation.

The sample is a subset of the population. Therefore, the sample correlation (r) will likely differ from the population correlation (*rho*).

Population

N = 100,000

id	weight	height
1	139	63
2	148	66
3	121	62
4	161	71
5	143	66
6	133	62
7	137	67
8	152	68
9	144	67
10	144	64
99991	168	65
99992	119	57
99993	136	66
99994	157	68
99995	142	64
99996	111	59
99997	136	61
99998	136	67
99999	153	68
100000	167	67

$$rho = ?$$

Sample

n = 10

id	weight	height
6157	127	63
95223	163	63
63567	142	66
12927	164	66
86418	108	61
86413	125	61
30903	126	65
49726	155	67
17403	125	64
30148	113	61

$$r = .680$$

NHST

NHST is a process that is typically used to determine if it is reasonable to state a population-level effect may not be exactly zero. To do so, NHST uses an index called the *p*-value.

Population
N = 100,000

id	weight	height
1	139	63
2	148	66
3	121	62
4	161	71
5	143	66
6	133	62
7	137	67
8	152	68
9	144	67
10	144	64
99991	168	65
99992	119	57
99993	136	66
99994	157	68
99995	142	64
99996	111	59
99997	136	61
99998	136	67
99999	153	68
100000	167	67

NHST: Study r = 0.680

- We begin by assuming the population correlation is exactly zero (i.e., *rho* = 0.000). We refer to this assumption, that the population correlation is exactly zero, as the null hypothesis.

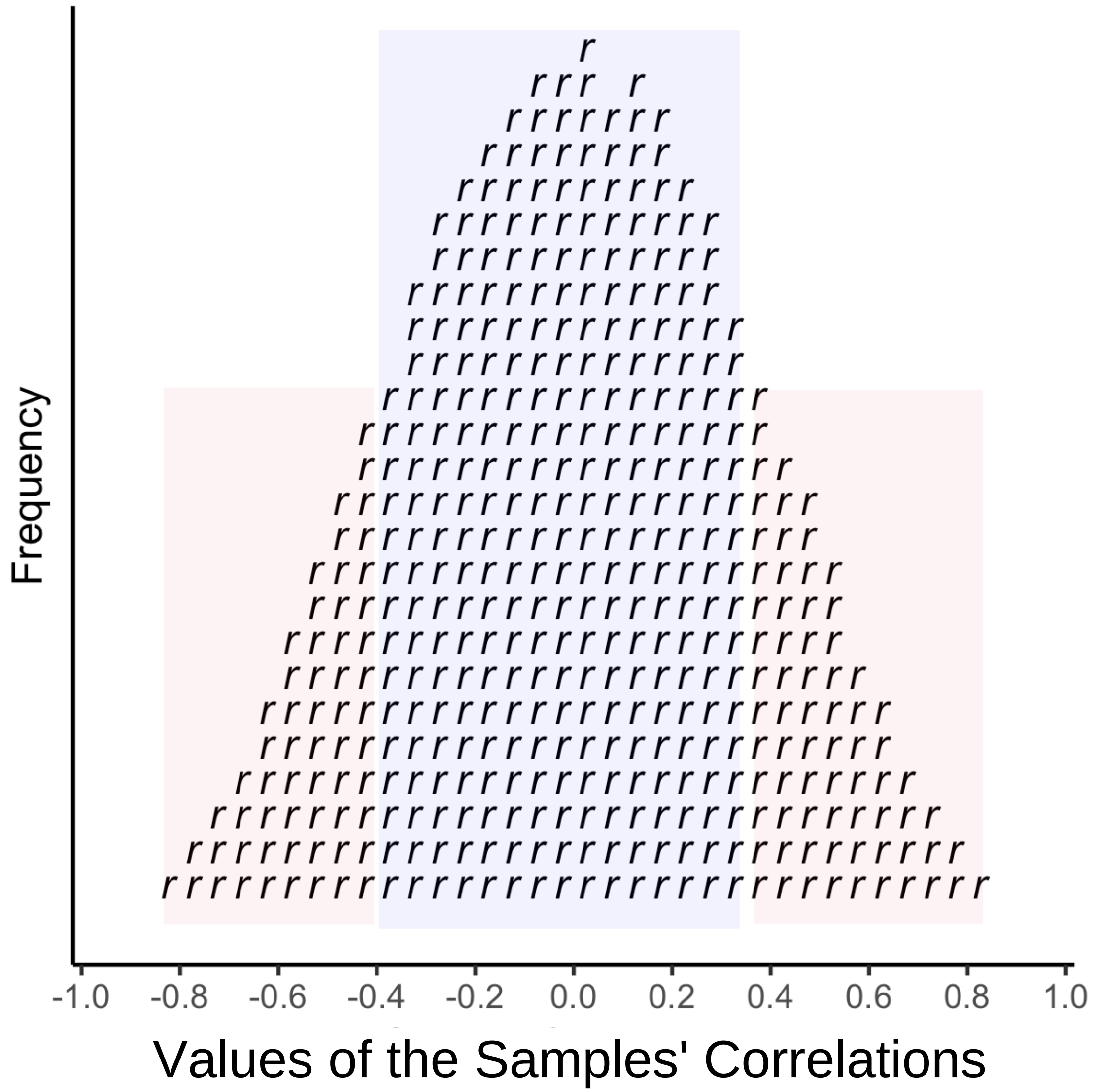
Population
N = 100,000

id	weight	height
1	139	63
2	148	66
3	121	62
4	161	71
5	143	66
6	133	62
7	137	67
8	152	68
9	144	67
10	144	64
99991	168	65
99992	119	57
99993	136	66
99994	157	68
99995	142	64
99996	111	59
99997	136	61
99998	136	67
99999	153	68
100000	167	67

Assume *rho = 0*

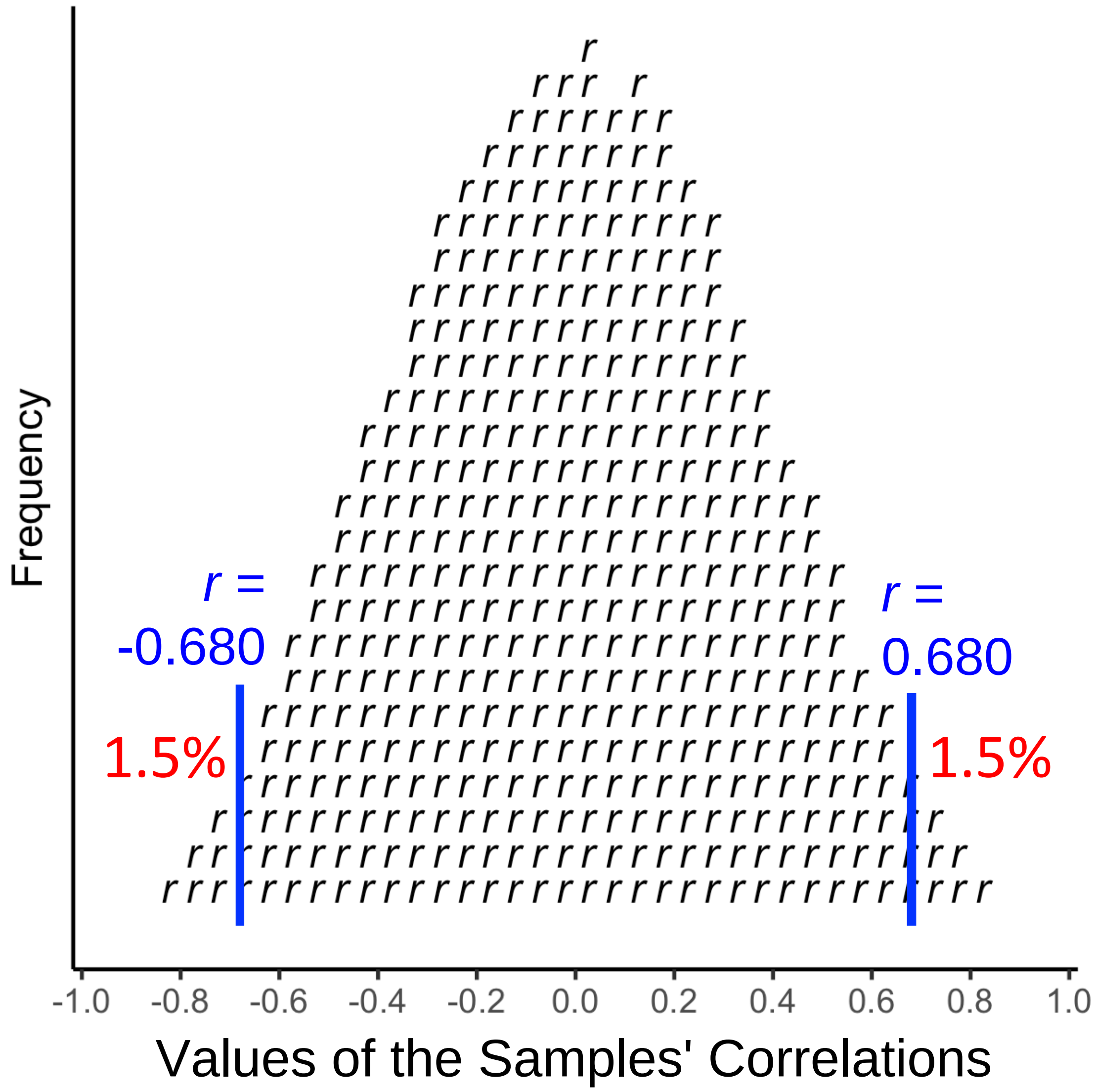
NHST: Study r = 0.680

- Next, imagine repeatedly taking samples (n = 10), from a population and obtaining a sample correlation for each. These imaginary correlations are illustrated here.
- Imaginary sample correlations whose value are close to zero are common in the distribution and are centered in the middle of the normal distribution (e.g., shaded in blue in the graph).
- Imaginary sample correlations whose values aren't close to zero are less common and are distributed toward the extreme ends of the distribution (e.g., shaded in red in the graph).



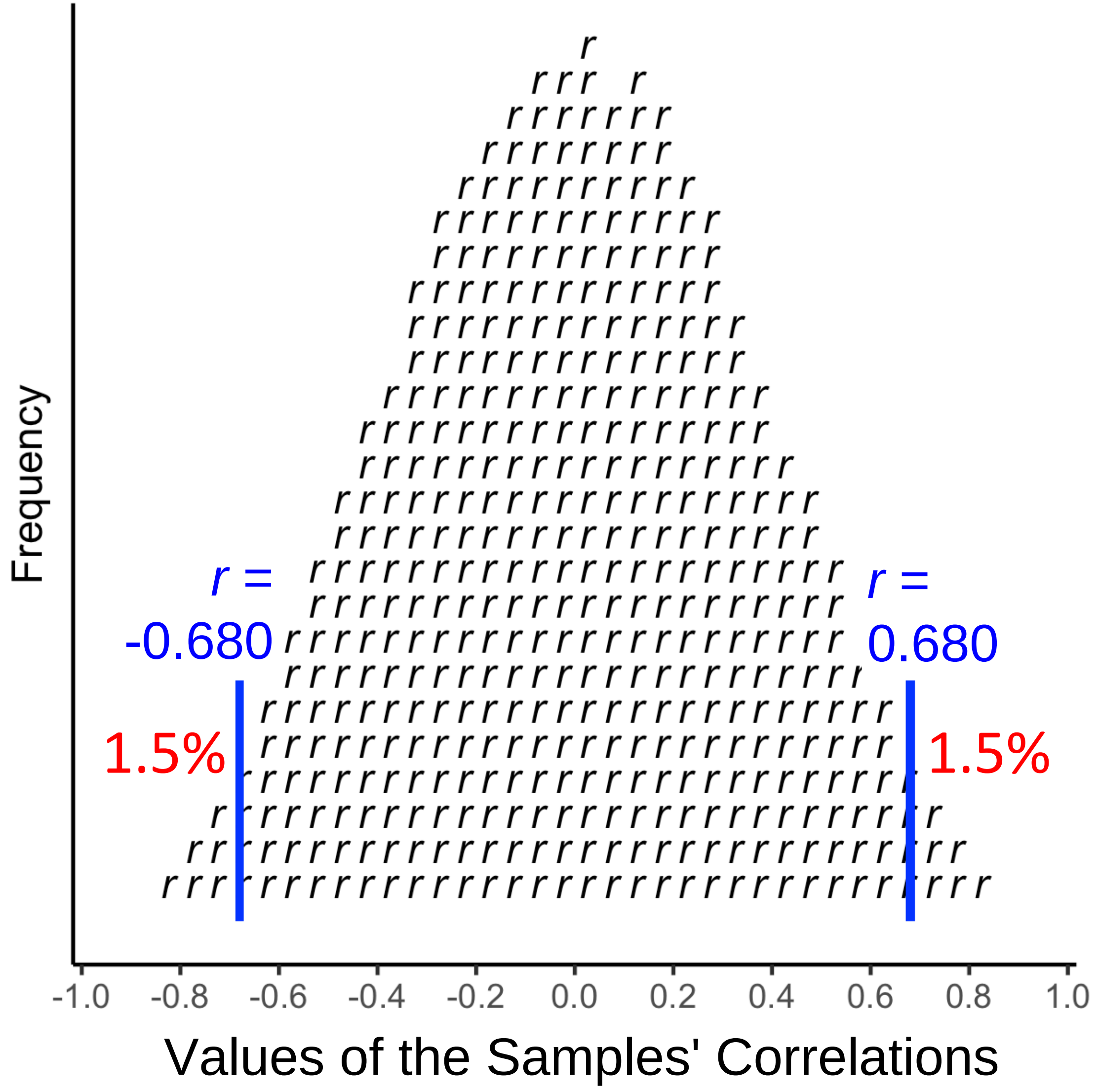
NHST: Study r = 0.680

- We use this set of imaginary correlations to help us interpret the study correlation (r = 0.680).
- We examine the percentage of imaginary correlations that are as extreme, or more extreme, **in either direction,** than the study correlation of r = 0.680.
- We see that 3% (1.5% + 1.5%) of the imaginary correlations are as extreme, or more extreme, than our study correlation.
- Therefore, p = .030 (two-tailed; with .150 or 1.5% on each tail side)



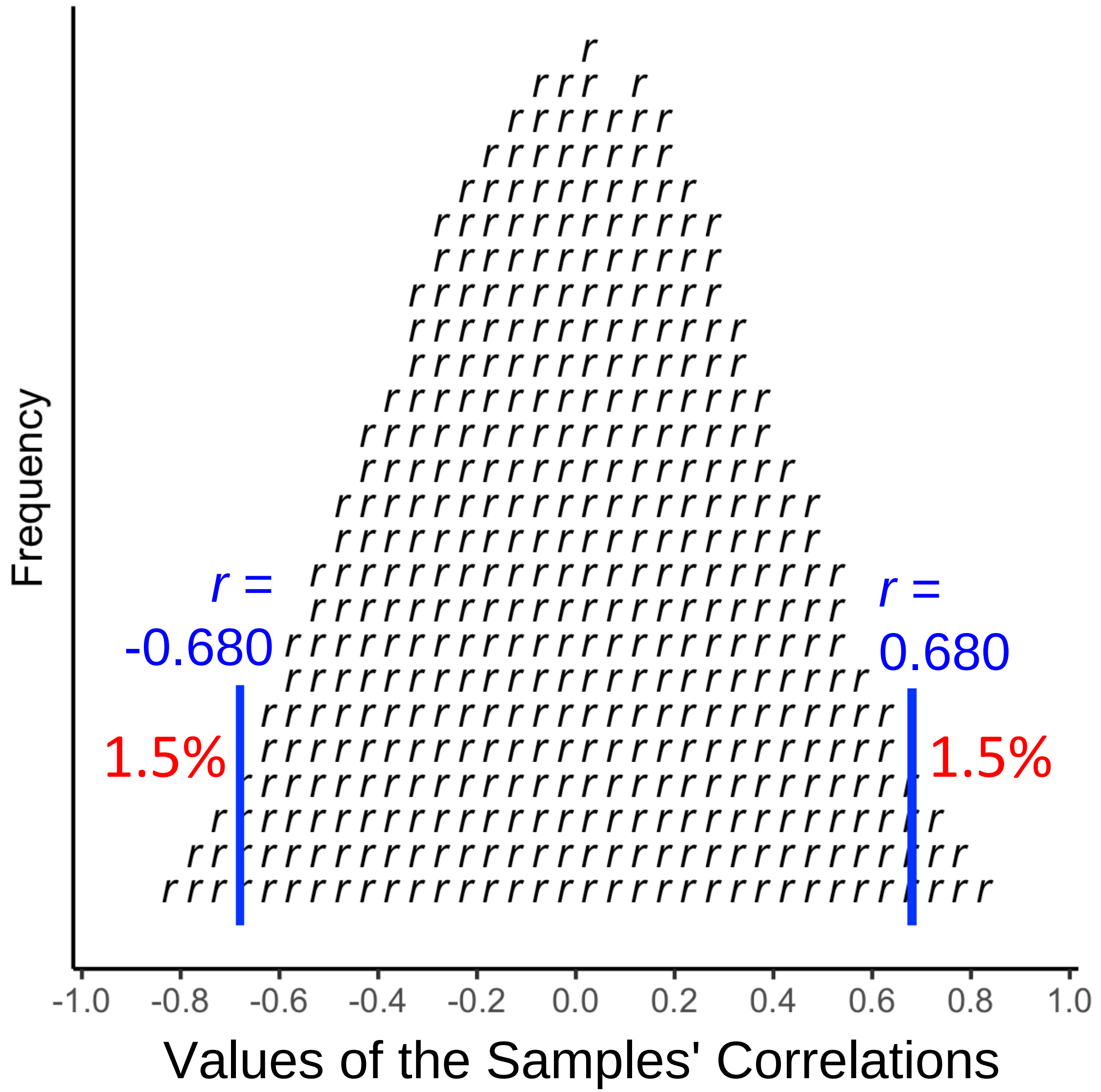
NHST: Study r = 0.680

- *p* = .030 means that, assuming the null hypothesis is true (i.e.,. *rho* = 0), there is a 3% probability of obtaining a study correlation as extreme, or more extreme, than our study's sample correlation.
- Therefore, our study's sample correlation is unlikely when starting with the assumption that the null hypothesis is true.
- If the *p*-value is below our cutoff, which for many studies is *p* < .050, we can reject our starting assumption, meaning, we can reject the null hypothesis.



NHST: Study r = 0.680

- When we reject the null hypothesis, we are concluding that the population correlation may not be zero.
- We subsequently refer to our study's correlation as "statistically significant."
- Of course, we may still be wrong. Our study correlation is still possible if the population correlation is zero; it is just unlikely!



American Statistical Association (ASA)

The use of *p*-values is associated with controversy. In 2019, the ASA issued some strong guidance about the use of *p*-values. The list below is directly quoted:

- "Don't base your conclusions solely on whether an association or effect was found to be "statistically significant" (i.e., the *p*-value passed some arbitrary threshold such as $p < 0.050$).
- Don't believe that an association or effect exists just because it was statistically significant.
- Don't believe that an association or effect is absent just because it was not statistically significant.
- Don't believe that your *p*-value gives the probability that chance alone produced the observed association or effect or the probability that your test hypothesis is true.
- Don't conclude anything about scientific or practical importance based on statistical significance (or lack thereof)." p. 1

THE AMERICAN STATISTICIAN
2019, VOL. 73, NO. S1, 1–19: Editorial
https://doi.org/10.1080/00031305.2019.1583913