

Introduction to hypothesis tests

Stat 250

[Click here for PDF version](#)

Do reward systems work?

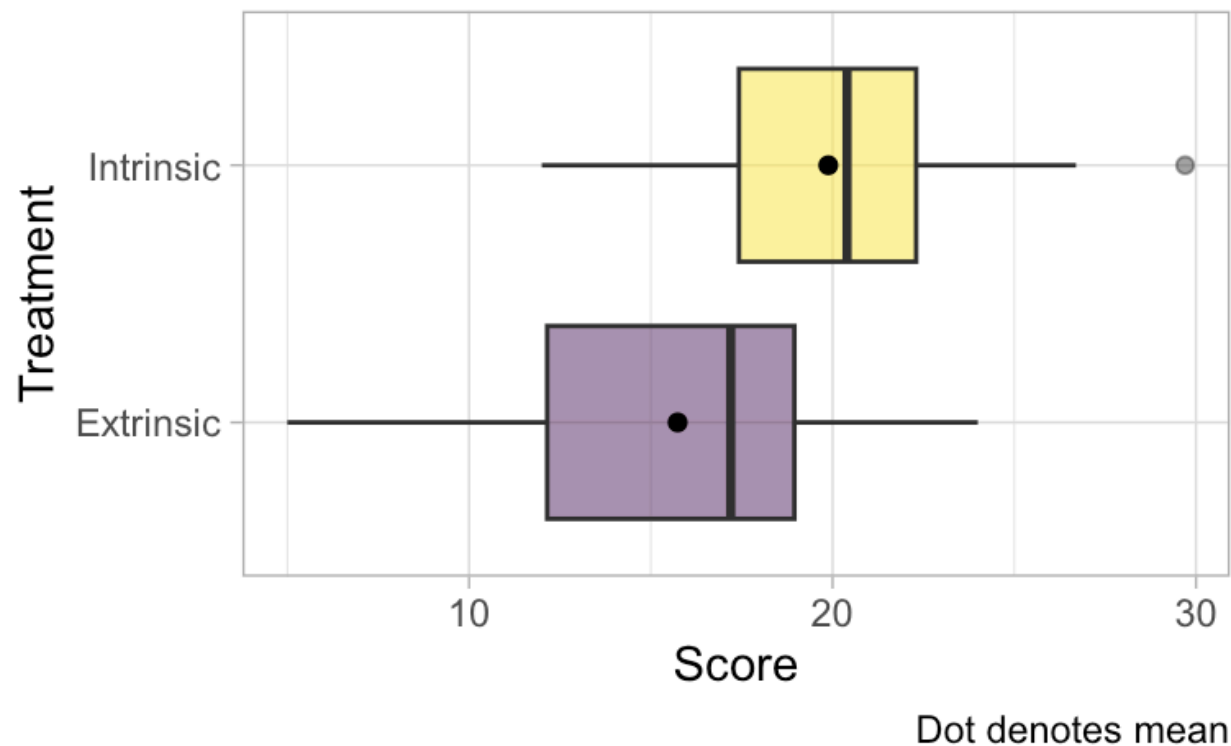
Evidence suggests that reward systems may operate in the opposite way from what is intended, e.g.,

- Ranking systems may decrease productivity;
- Rewards may not stimulate learning

Experiment

- 47 subjects with considerable experience in creativity were recruited
- Randomly assigned to either intrinsic- or extrinsic-motivation group
- Subjects completed a questionnaire related to either intrinsic or extrinsic reasons for writing
- All subjects were asked to write a Haiku about laughter
- Poems were scored by a panel of 12 poets, evaluated on 40-point creativity scale

Results



Treatment	min	Q1	median	Q3	max	mean
Extrinsic	5	12.2	17.2	19.0	24.0	15.7
Intrinsic	12	17.4	20.4	22.3	29.7	19.9

The logic of hypothesis testing

1. Formulate two competing **hypotheses** about the population
2. Calculate a **test statistic** summarizing the relevant information to the claims
3. Look at the **behavior of the test statistic** assuming that the initial claim is true
4. **Compare** the observed test statistic to the expected behavior (the distribution created in step 3).

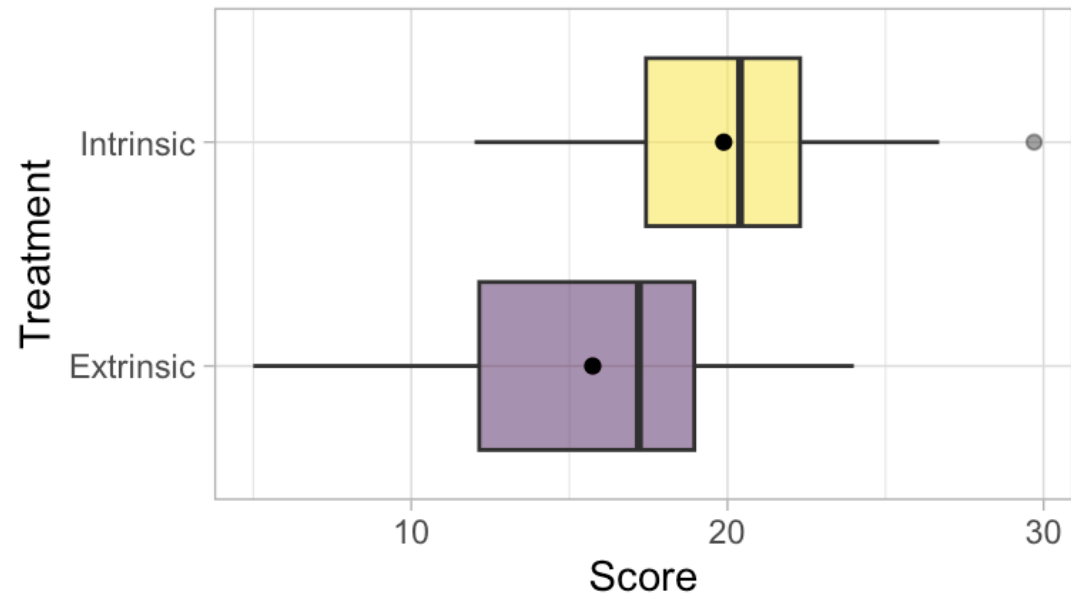
How likely is it that the observed test statistic occurred by chance alone?

Your turn

Write down the competing hypotheses being tested by the researchers, both in words and using proper notation.

Permutation resampling

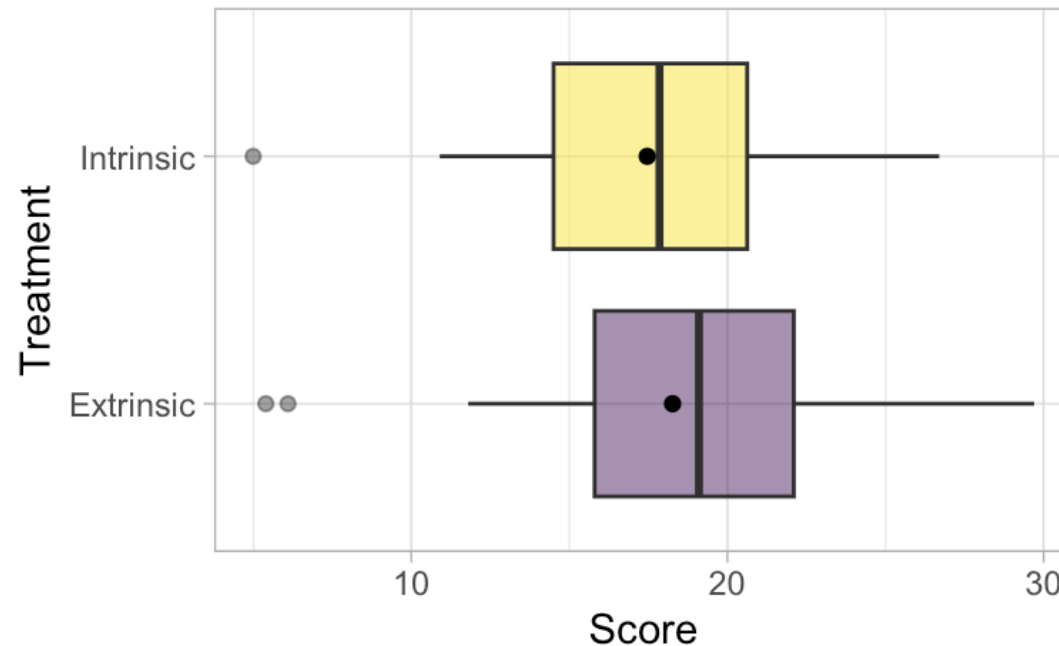
Intrinsic						Extrinsic					
12	16.6	19.1	20.5	22.1	24	5	11.8	15	17.4	18.7	21.2
12	17.2	19.3	20.6	22.2	24.3	5.4	12	16.8	17.5	19.2	22.1
12.9	17.5	19.8	21.3	22.6	26.7	6.1	12.3	17.2	18.5	19.5	24
13.6	18.2	20.3	21.6	23.1	29.7	10.9	14.8	17.2	18.7	20.7	



Dot denotes mean

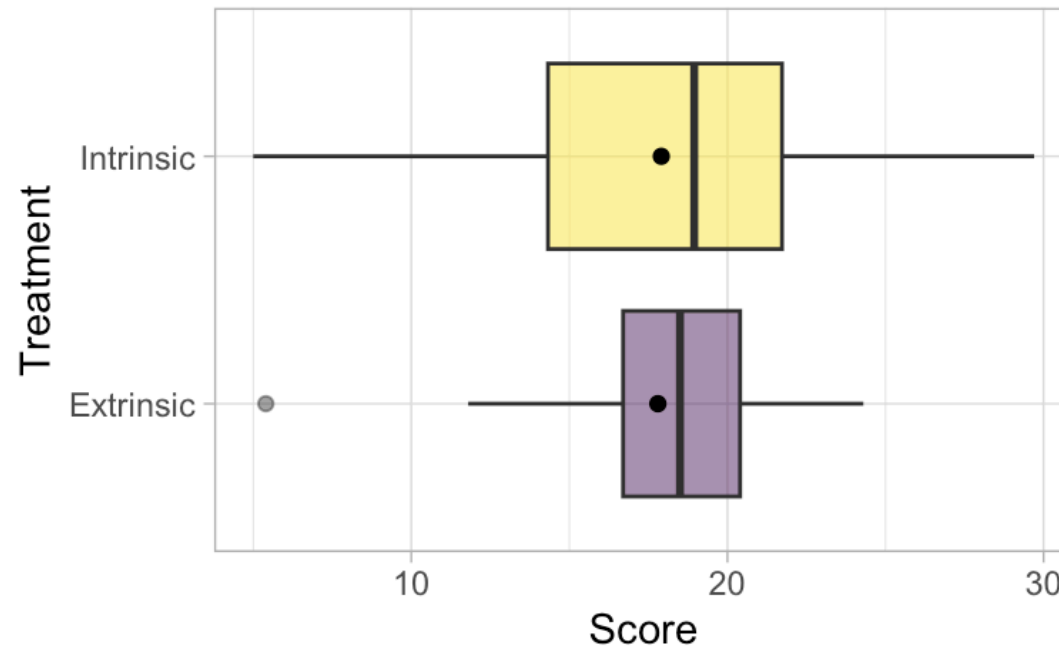
Permutation resampling

Intrinsic						Extrinsic					
5	12.9	17.2	18.2	20.3	21.3	5.4	12	17.5	19.3	22.1	24
10.9	13.6	17.2	18.7	20.6	22.2	6.1	15	18.5	19.8	22.1	24.3
12	14.8	17.2	19.2	20.7	23.1	11.8	16.6	18.7	20.5	22.6	29.7
12.3	16.8	17.5	19.5	21.2	26.7	12	17.4	19.1	21.6	24	



Permutation resampling

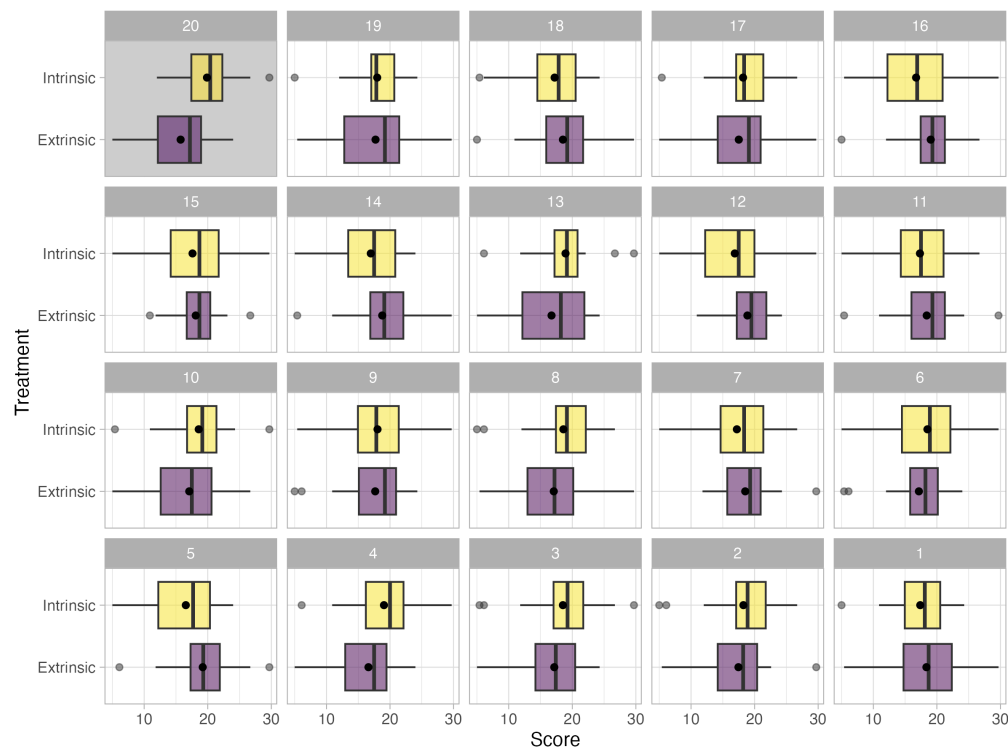
Intrinsic						Extrinsic					
5	12	17.2	19.2	21.3	22.6	5.4	13.6	17.2	18.7	20.3	24
6.1	12.9	17.5	19.8	21.6	23.1	11.8	16.6	17.4	19.1	20.5	24
10.9	14.8	17.5	20.7	22.1	26.7	12	16.8	18.2	19.3	20.6	24.3
12	15	18.7	21.2	22.2	29.7	12.3	17.2	18.5	19.5	22.1	



Your turn

How likely is it that the observed results occurred by chance alone?

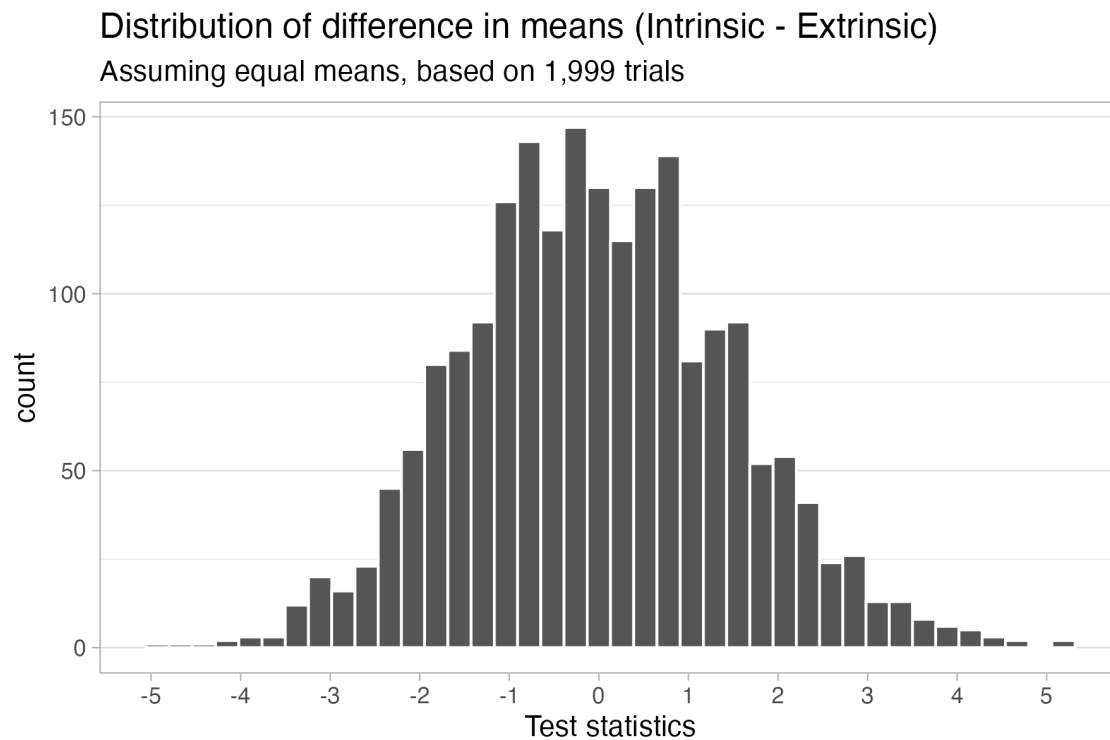
What does this say about the null hypothesis?



Your turn

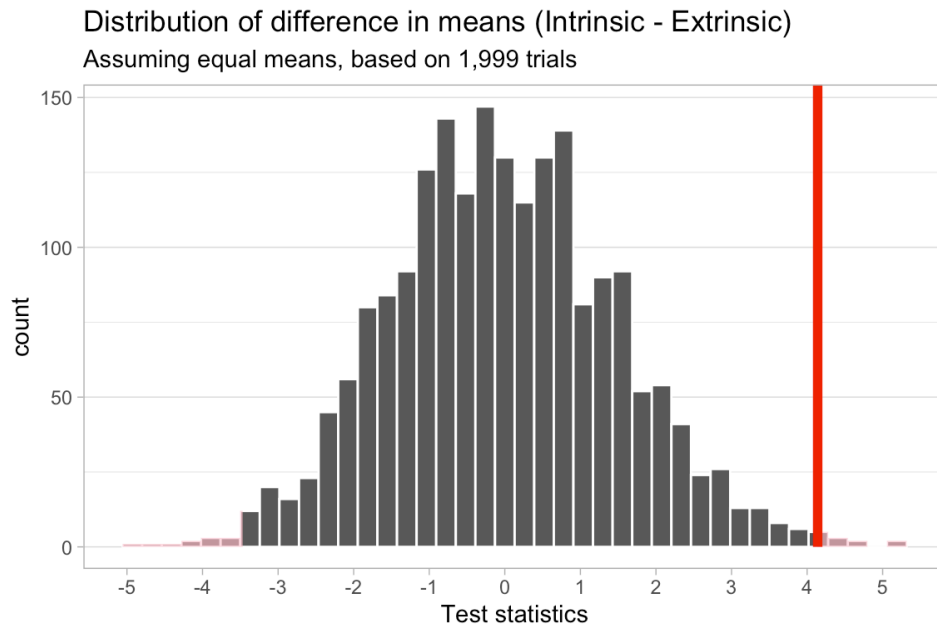
How likely is it that the observed test statistic occurred by chance alone?

What does this say about the null hypothesis?



p -value

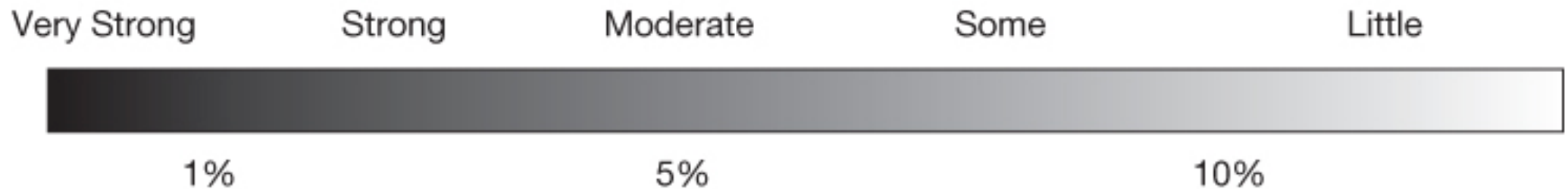
Definition: fraction of times the random test statistic exceeds the original test statistic



Observed test statistic:
10 of the random test
statistics exceed 4.2

Strength of evidence

p-values provide a *continuous* measurement of the strength of evidence against the null hypothesis



 A bit more on the null distribution...

Permutation test algorithm

1. Pool the data values
2. Draw a resample of size without replacement, assign these values to group 1. Assign the remaining values to group 2.
3. Calculate the test statistic comparing the samples from the resampled groups.
4. Repeat steps 2 and 3 until we have enough samples.
5. Estimate the p-value as the proportion of times the observed test statistic exceeds the original (observed) test statistic

Your turn

Work through the additional examples with a neighbor, ask questions as they come up