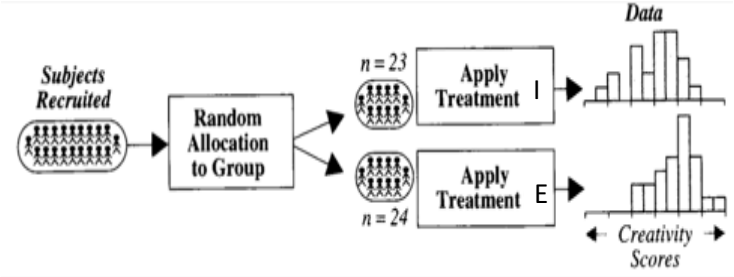
Drawing Statistical Conclusions

MEASURING UNCERTAINTY IN RANDOMIZED AND OBSERVATIONAL STUDIES

Quantifying Uncertainty

Creativity Study

I



→ Population mean: μ_I

→ Population mean: μ_E

- If the questionnaires had no effect, then we would expect:

$$\mu_I = \mu_E \leftrightarrow \mu_I - \mu_E = 0$$ **(NULL HYPOTHESIS)**

- We have discussed that the sample means $\bar{Y}_I$ and $\bar{Y}_E$ are good estimates of μ_I, μ_E

→ $\bar{Y}_I - \bar{Y}_E$ is a reasonable estimate of $\mu_I - \mu_E$ **(TEST STATISTIC)**

- We can compute this **OBSERVED DIFFERENCE** in sample means: 4.14420
- Is 4.14420 large enough for us to conclude that $\mu_I \neq \mu_E$? **(ALTERNATE HYPOTHESIS)**

Creativity Study

- To quantify "large", we can re-randomly allocate units to two groups and re-compute the difference in sample means many times
- We say that a re-computed difference is **MORE EXTREME** provided

$$|\text{re-computed difference}| > \text{abs}|\bar{Y}_I - \bar{Y}_E|$$

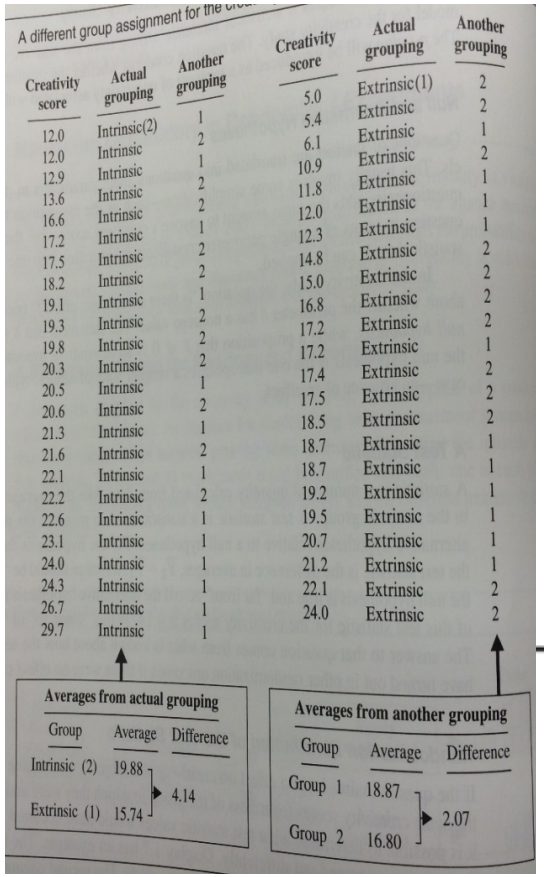
- Suppose that $\frac{number\ of\ more\ extreme\ re{-}computed\ differences}{total\ number\ of\ re{-}random\ allocations} = pval$ **(P-VALUE)**
- If *pval* is very small (say 0.01), this provides evidence that the intrinsic/extrinsic group result would be very unusual if the questionnaire had no effect
- If *pval* is very big (say 0.2), this provides no evidence that the intrinsic/extrinsic group result would be very unusual if the questionnaire had no effect

Creativity Study: Testing the Hypothesis

Number of random regroupings:1.6 x 10^{13}

Half a year with a computer that can perform a million calculations per second!

(Still only 84% of the U.S. Federal Government debt in dollars, though)



A different group assignment for the

Creativity score	Actual grouping	Another grouping
12.0	Intrinsic(2)	1
12.0	Intrinsic	2
12.9	Intrinsic	1
13.6	Intrinsic	2
16.6	Intrinsic	2
17.2	Intrinsic	1
17.5	Intrinsic	2
18.2	Intrinsic	2
19.1	Intrinsic	1
19.3	Intrinsic	2
19.8	Intrinsic	2
20.3	Intrinsic	2
20.5	Intrinsic	1
20.6	Intrinsic	2
21.3	Intrinsic	1
21.6	Intrinsic	2
22.1	Intrinsic	1
22.2	Intrinsic	2
22.6	Intrinsic	1
23.1	Intrinsic	1
24.0	Intrinsic	1
24.3	Intrinsic	1
26.7	Intrinsic	1
29.7	Intrinsic	1
5.0	Extrinsic(1)	2
5.4	Extrinsic	2
6.1	Extrinsic	1
10.9	Extrinsic	2
11.8	Extrinsic	1
12.0	Extrinsic	1
12.3	Extrinsic	1
14.8	Extrinsic	2
15.0	Extrinsic	2
16.8	Extrinsic	2
17.2	Extrinsic	2
17.2	Extrinsic	1
17.4	Extrinsic	2
17.5	Extrinsic	2
18.5	Extrinsic	2
18.7	Extrinsic	1
18.7	Extrinsic	1
19.2	Extrinsic	1
19.5	Extrinsic	1
20.7	Extrinsic	1
21.2	Extrinsic	1
22.1	Extrinsic	2
24.0	Extrinsic	2

Averages from actual grouping

Group	Average	Difference
Intrinsic (2)	19.88	4.14
Extrinsic (1)	15.74	

Averages from another grouping

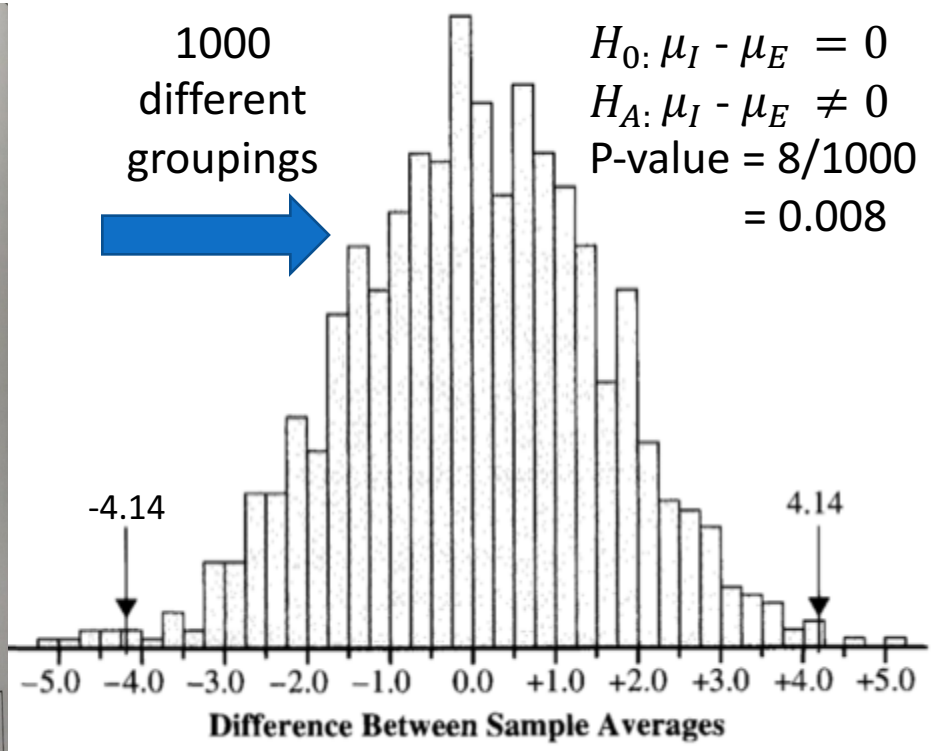
Group	Average	Difference
Group 1	18.87	2.07
Group 2	16.80	

1000 different groupings

$H_{0:}\ \mu_I - \mu_E = 0$

$H_{A:}\ \mu_I - \mu_E \neq 0$

P-value = 8/1000 = 0.008



From Randomized to Observational Studies

- In the Creativity study, the Intrinsic/Extrinsic groups were randomly assigned to subjects
- This motivated comparing the observed difference to re-randomized difference to test a hypothesis about the questionnaire having no effect.
- This is known as a **RANDOMIZATION TEST**
- In observational studies, the groups are not randomly assigned
- Though not technically the same test, we can still apply exactly the same re-randomization idea to observational data
- However, now it is called a **PERMUTATION TEST**

Age Discrimination

In the United States, it is illegal to discriminate against people based on various attributes. One example is age. An active lawsuit, filed August 30, 2011, in the Las Angeles District Office is a case against the American Somoa Government for systematic age discrimination by preferentially firing older workers.

Is there evidence for age discrimination in this study?