Review of MATH 300 concepts (t-tests, sampling distributions, standard error, central limit theorem)
Resources: creative.sav

Do grading systems promote creativity in students? Do rewards and praise stimulate children to learn? Although reward systems are deeply imbedded in schools and in the workplace, a growing body of evidence suggests that rewards may operate in precisely the opposite way from what is intended.

Psychologist Teresa Amabile demonstrated this phenomenon in an experiment concerning the effects of intrinsic and extrinsic motivation on creativity. Subjects with considerable experience in creative writing were randomly assigned to one of two treatment groups:

1) 24 of the subjects were placed in the "intrinsic" treatment group
2) 23 of the subjects were placed in the "extrinsic" treatment group

Those in the "intrinsic" group completed a questionnaire that involved ranking intrinsic reasons for writing (see #1 below). The questionnaire was intended as a device to establish a thought pattern concerning intrinsic motivation – doing something because doing it brings satisfaction. The subjects in the "extrinsic" group completed a questionnaire that got them thinking about extrinsic motivation (see #2 below) – doing something because a reward is associated with its completion.

#1: Rank the following list of reasons for writing, in order of personal importance to you
___ You enjoy the opportunity for self-expression
___ You achieve new insights through your writing
___ You feel relaxed when writing
___ etc

#2: Rank the following list of reasons for writing, in order of personal importance to you
___ You want your teachers to be impressed with your writing
___ You enjoy public recognition of your work
___ You know that the best jobs require good writing skills
___ etc

After completing the questionnaire, all subjects were asked to write a poem in the Haiku style about "laughter." All poems were submitted to 12 poets who evaluated them on a 40-point scale of creativity. Judges were not told of the purpose of the study. The average ratings given by the 12 judges are shown below.

Intrinsic Group (n = 24)		Extrinsic Group (n = 23)	
12.0	20.5	5.0	17.4
12.0	20.6	5.4	17.5
12.9	21.3	6.1	18.5
13.6	21.6	10.9	18.7
16.6	22.1	11.8	18.7
17.2	22.2	12.0	19.2
17.5	22.6	12.3	19.5
18.2	23.1	14.8	20.7
19.1	24.0	15.0	21.2
19.3	24.3	16.8	22.1
19.8	26.7	17.2	24.0
20.3	29.7	17.2	

Minimum Rating	12.0	5.0
1st quartile	17.25	12.0
Median	20.4	17.2
3rd quartile	22.5	19.2
Maximum	29.7	24.0
Mean Rating	19.88	15.74
Standard Deviation	4.44	5.25

Is there any evidence that creativity scores tend to be affected by the type of motivation (intrinsic or extrinsic) induced by the questionnaires?

Source: Amabile, T. "Motivation and Creativity: Effects of Motivational Orientation on Creative Writers," *Journal of Personality and Social Psychologyi 48(2) 1985*

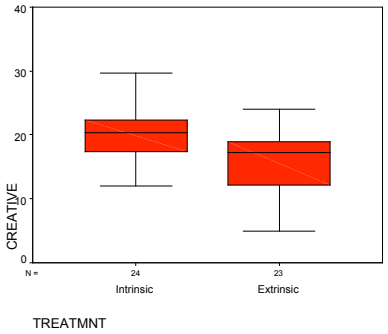
1) State the null and alternative hypotheses in this study. Which hypothesis do we assume is true?

H_0: $\mu_{intrinsic} = \mu_{extrinsic}$

H_0: $\mu_{intrinsic} \neq \mu_{extrinsic}$

We proceed under the assumption that the null hypothesis is true. In this example, we assume the treatments have no impact on creativity ratings (i.e. the difference between groups is due to random error)

2) Create side-by-side boxplots to compare the two distributions of creativity ratings. Which treatment appears to yield higher creativity scores?



The intrinsic group received higher ratings. We must conduct a statistical test of this hypothesis to see if the difference in ratings is actually due to a treatment effect.

3) Let's compare these differences using a back-to-back stemplot. Which plot do you prefer: stemplots or boxplots?

Intrinsic	Stem	Extrinsic
	5	04
	6	1
	7	
	8	
	9	
	10	9
	11	8
900	12	03
6	13	
	14	8
	15	0
6	16	8
52	17	2245
2	18	577
831	19	25
653	20	7
63	21	2
621	22	1
1	23	
30	24	0
	25	
7	26	
	27	
	28	
7	29	

The stemplot lets you see the actual distribution of all the scores.

Stemplots also let you assess normality.

4) What type of hypothesis test will we need to conduct in order to answer the research question? What assumptions must be met in order to use this hypothesis test?

An independent samples t-test will answer the research question.

Assumptions:

1) Independence
   We'll have to assume the raters independently scored the haikus

2) Normality
   The stemplot shows the data is roughly normal.

5) Before we conduct our hypothesis test, we need to assess the possible decision errors. What are the potential decision errors in this study? What significance level will you choose for your analysis? What does this significance level represent?

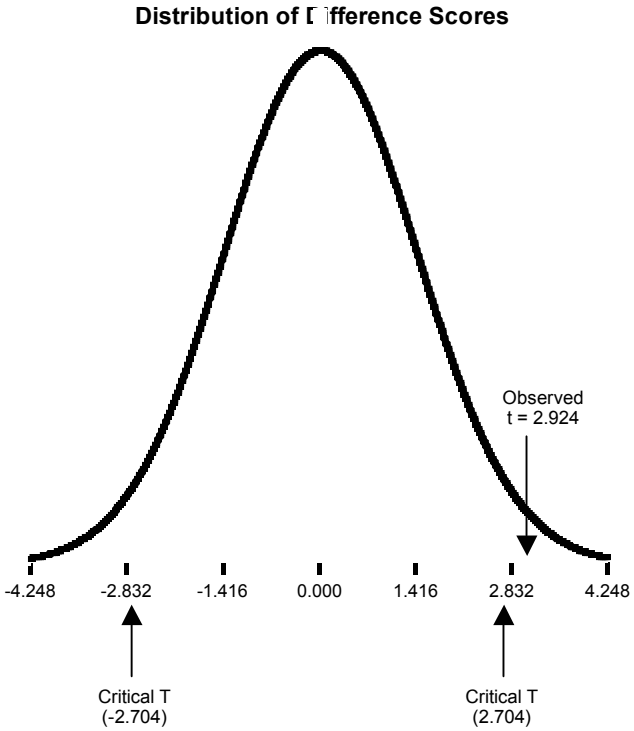
		Decision (based on observed sample)	
		Fail to reject H_0	*Reject H_0*
Reality	*H_0 is true*	Correct Decision	Type I Error (α)
	H_0 is false	Type II Error (ß)	Correct Decision

		Decision (based on observed sample)	
		Treatments have no effect	*Intrinsic Motivation increases (or decreases) creativity*
Reality	*No treatment effect (intrinsic motivation and extrinsic motivation have the same effect on creativity)*	We correctly conclude the treatments have impact on creativity	We incorrectly conclude that intrinsic motivation increases (or decreases) creativity Schools could spend lots of money getting rid of traditional grades with the mistaken belief that it will increase achievement.
	Treatment effect exists (intrinsic motivation and extrinsic motivation have the same effect on creativity)	We incorrectly conclude that intrinsic and extrinsic motivation yield the same creativity scores. Student achievement is not increased, because schools stay with a mix of extrinsic and intrinsic motivational techniques.	We correctly conclude that intrinsic motivation increases (or decreases) creativity

In my opinion, the Type I error is most costly. I'm going to choose a relatively low alpha level of 0.01 (which represents the probability of making a Type I error).

6) What test statistic do we use to conduct this analysis? How is this test statistic distributed (what is its sampling distribution)? How do you know its distribution? What is the standard error of this test statistic? Calculate the value of the test statistic.

7) Sketch the distribution of this test statistic. Using $\alpha = 0.01$, find the critical region of this distribution. Finally, locate your observed test statistic on this distribution. Do you reject or retain the null hypothesis?



$$t_{n_1+n_2-2} = \frac{Point\ Estimate - Hypothesized}{Standard\ Error}$$

$$= \frac{\left(\overline{X}_I - \overline{X}_E\right) - 0}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}} =$$

$$= \frac{19.88 - 15.74}{1.416} = 2.924 = t_{45}$$

This is a two-sided test, so we need to find a value of Z such that 0.5% of the area under the curve is to the right of Z (and 0.5% is to the left of –Z).

Using the table of critical t-values (with 45 df) in our textbook, we find that $t_{.005} = -2.704$

Since our observed t-value falls in this critical region, we conclude that the null hypothesis must be false (the mean of the differences is probably not zero).

8) Calculate the p-value of your test statistic. Interpret this p-value.

p-value: We cannot compute it directly, but we know it's a bit less than 0.01.

If the null hypothesis were true – if the type of motivation had no impact on creativity – then it would be extremely unusual to find our observed difference between groups.

Stated other ways…

"A p-value is the probability that random error alone leads to a test statistic as extreme or more extreme than the one observed." – Statistical Sleuth

"A p-value of a test is the chance of getting a big test statistic – assuming the null hypothesis to be correct. P is not the chance of the null hypothesis being correct." -- Freedman

9) Suppose we constructed a 99% confidence interval from our sample data. Would zero be contained in this interval? How do you know? Construct the confidence interval to verify your answers. Interpret the confidence interval.

Zero will not be in the interval, since we rejected the null hypothesis at an alpha level of 0.01.

The confidence interval is calculated to be: 0.311 to 7.969

We're 99% sure the mean difference score would be contained within this interval (if we repeatedly took samples and calculated confidence intervals with the same width, 99% of them would contain the true population mean difference)

It looks as though the intrinsic group scores 1-8 points higher than the extrinsic group.

$$\text{Point Estimate} \pm t_{cv} SE$$
$$= 4.14 \pm t_{45,.005}(1.416) = 4.14 \pm 2.704(1.416) = (0.311, 7.969)$$

10) Several websites, such as http://calculators.stat.ucla.edu/powercalc/ allow you to calculate the power of your test if you make several assumptions about your population. If we assume the true population mean for the intrinsic group is 20 and the true population mean for the extrinsic group is 16, the power of our test is found to be 0.5541. Interpret this value for the power of our test. What is the value of β for our study? Interpret β.

Power refers to the chances we will reject the null hypothesis if, in fact, the null hypothesis is false. In other words, it's the probability that we find significance when significance exists.

We had a 55% chance of finding a real significant difference. We could have increased the power of this study by increasing our sample size or by lowering our alpha level.

Beta, which is equal to (1 – Power), equals .45 in this study. This refers to the probability of making a Type II error (incorrectly retaining a false null hypothesis).

11) What conclusions can you make based on your analysis of this study? Does intrinsic motivation yield higher creativity in students?

This is an experimental study – we randomly assigned subjects to treatments. Therefore, we can conclude causality.

It appears as though intrinsic motivation yields higher creativity scores. We must be careful to ensure our sample is representative of our target population. We must also be careful to conclude that intrinsic motivation increases creativity **as measured through haikus**.

12) In this study, the poems were given to the judges in random order. Why was this important?

We had to ensure the judges' ratings were independent. Some judges may give higher ratings to the first several haikus they see. Other judges might give higher ratings to the last several they read.