4/26/2022

Continue with hypothesis testing

Example.

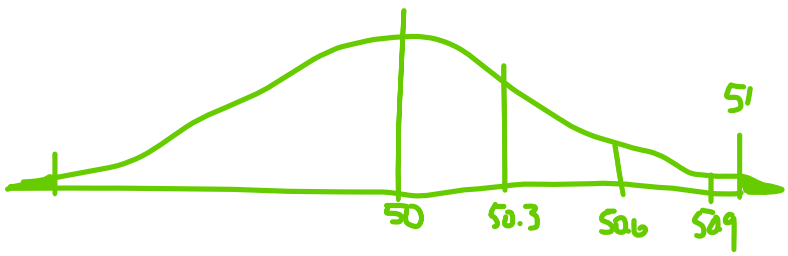
Suppose that we want to test the hypothesis with a significance level of $\alpha = 0.01$ that the climate has changed since industrialization. Suppose that the mean temperature throughout history is 50 degrees. During the last 40 years, the mean temperature has been 51 degrees and suppose the population standard deviation is 2 degrees. What can we conclude?

(from Worksheet #3).

One sample test of means. Two-tailed test (is this data different than the past?)

$$H_0: \mu = 50$$
$$H_a: \mu \neq 50$$

Go to excel to calculate the p-value
We will calculate with the normal distribution, and compare with the t-distribution. Some changes in the calculation procedure are needed.



hypothesis test example (climate)		
mu_0	50	
mean_samp	51	
st.dev	2	
sample size	40	
st.error	0.316228	
p-value	0.000783	one-tailed test
	0.001565	two-tailed
using the normal distribution		
alpha	0.01	

Since 0.001565<0.01, we reject the null hypothesis and conclude the climate of the last 40 years is different from the previous climate.

There IS sufficient evidence to conclude that the climate in the last 40 year is different from the previous climate records.

What if we use the t-distribution.

to use t-distribution: one additional step before calculating the p-value

t-score	3.162278	
p-value	0.001514	one-tailed
	0.001514	one-tailed
	0.003027	two-tailed
	0.003027	two-tailed

The p-value for the two-tailed test is still less than alpha, and we still reject the null hypothesis.

Example.
The NCHS report indicated that in 2002 the prevalence of cigarette smoking among American adults was 21.1%. Data on prevalent smoking in n=3,536 participants who attended the seventh examination of the Offspring in the Framingham Heart Study indicated that 482 of the respondents were currently smoking at the time of the exam. Suppose we want to assess whether the prevalence of smoking is different in the Framingham Offspring sample given the focus on cardiovascular health in that community. Is there evidence of a statistically different prevalence of smoking in the Framingham Offspring study as compared to the prevalence among all Americans?
(Worksheet #6)

Test of proportions. Two-tailed test (different).

$$H_0: p = 21.1\%$$
$$H_a: p \neq 21.1\%$$

Calculations in Excel

proportion test (smoking)	
p_0	21.10%
p_sample	13.63%
count	482
sample size	3536
st.error	0.006861572
p-value	6.79954E-28

one-tailed p-value, double to get the two-tailed value: this won't make any practical difference here, but it can sometimes.
p-value<<alpha. So reject the null hypothesis.
That means there IS sufficient evidence to conclude that the study participants smoke less often than the general public.

Only need to use the normal distribution for proportions. Never t. (T is only for means).

Comparing two samples.
In general, the null hypothesis for comparing two samples is to assume they are identical (their means or proportions are the same).

$$H_0: \mu_1 = \mu_2$$

OR

$$H_0: \mu_1 - \mu_2 = 0$$

For proportions:

$$H_0: p_1 = p_2$$

OR

$$H_0: p_1 - p_2 = 0$$

Standard error calculation: for the difference of proportions.

$$\sigma_{p_1 - p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Example.
Time magazine reported the result of a telephone poll of 800 adult Americans. The question posed of the Americans who were surveyed was: "Should the federal tax on cigarettes be raised to pay for health care reform?" The results of the survey were:

	Non-Smokers	Smokers
Sample size	605	195
“yes”	351	41

Is there sufficient evidence at the α = 0.05 level, say, to conclude that the two populations — smokers and non-smokers — differ significantly with respect to their opinions?
(Worksheet #7)

This a two-sample proportion test. Two-tailed test.

$$H_0: p_1 = p_2$$
$$H_a: p_1 \neq p_2$$

Calculations in Excel.

	Non-Smokers	Smokers
Sample size	605	195
“yes”	351	41
p_samp	0.580165	0.210256
	col1	col2
H0	0	p1-p2=
p1-p2	0.369909	
st.error	0.035414	

p-value	0	so small, Excel cannot tell us what it is	
z-score	10.44535		two-tailed p-value
			0

This 0<<0.05, so we reject the null hypothesis. We conclude that there is very strong evidence to support the claim that the two groups (smokers and non-smokers) disagree about whether taxes on cigarettes should be raised to pay for health care reform.

Two-Sample T-Test
These tests of means.

Paired tests (the data is dependent—values are paired with one or more factors)
Independent tests: can be broken down a little further

Pooled (one standard deviation calculation for both groups combined)

Unpooled (calculate the st.deviations separately, and then combine them in a kinda complicated way)—is the better way, but a lot will estimate some of the complexity.

St.error formulas for the pooled and unpooled cases.

$$S_P^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{(n_1 - 1) + (n_2 - 1)}$$

This is the pooled variance. Take the square root of this to get the pooled standard deviation.

$$SE_{(x_1 - x_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

we will use this version when we have to do this by hand.

Degrees of freedom

$$df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

For the paired test: use n-1 as the degrees of freedom (after finding differences in the paired data, treat like a one-sample test).

For the independent: use n-2 as the degrees of freedom.
Some people will use n-1 and the n from the smaller sample size.

Example.
Within a school district, students were randomly assigned to one of two Math teachers - Ms. Smith and Mr. Jones. After the assignment, Ms. Smith had 30 students, and Mr. Jones had 25 students. At the end of the year, each class took the same standardized test. Ms. Smith's students had an average test score of 85, with a standard deviation of 15; and Mr. Jones' students had an average test score of 78, with a standard deviation of 10. Test the hypothesis that Ms. Smith and Mr. Jones are equally effective teachers. Use a 0.10 level of significance. (Assume that student performance is approximately normal.) (Worksheet #8)
Two-sample t-test. Test of means. Independent samples. (one clue about independence: they don't have to be the sample size. Dependent or paired samples must be the same size. Also, summaries of data must be treated as independent.)

Example.
Forty-four sixth graders were randomly selected from a school district. Then, they were divided into 22 matched pairs, each pair having equal IQ's. One member of each pair was randomly selected to receive special training. Then, all of the students were given an IQ test. Test results are summarized below.

Pairs	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Training	95	89	76	92	91	53	67	88	75	85	90
No Training	90	85	73	90	90	53	68	90	78	89	95
Pairs	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**
Training	85	87	85	85	68	81	84	71	46	75	80
No Training	83	83	83	83	65	79	83	60	47	77	83

Do these results provide evidence that the training changed student performance? Use an 0.05 level of significance. Assume that the mean differences are approximately normally distributed.
(Worksheet #10)

Two-sample t-test. Test of means. Paired t-test. One-tailed. ("pairs", "matched", and the groups are the same size, and we have raw data).

$$H_0: \mu_1 = \mu_2$$

Sometimes:

$$H_0: \delta = 0$$

In a paired test, we are going to calculate the differences between the pairs (each individual pair), and then test on those differences.

$$H_a: \mu_1 > \mu_2$$

$$H_a: \delta > 0$$

Calculate in Excel.

mu_0	0
mean_sample	0.954545
st.dev.	3.565406
sample size	22
st.error	0.760147
t-score	1.255738
p-value	0.111502

0.1115 > 0.05.
Fail to reject the null.

We can conclude that there is not enough evidence to think that the training made any difference.

Do an independent unpaired two sample test in Excel using data, and the Data Analysis tool pack.

See Excel.

H_0	mu_1=mu_2
H_a	mu_1 =/ mu_2
mean_1	72025
mean_2	67222.22
st.dev._1	9152.657
st.dev._2	10119.26
n_1	12
n_2	9
st.error	4284.699
df	19
t-score	1.120914
p-value	0.276293

t-Test: Two-Sample Assuming Unequal Variances

	Female Salary	*Male Salary*
Mean	72025	67222.22
Variance	83771136.36	1.02E+08
Observations	12	9
Hypothesized Mean Difference	0	
df	16	
t Stat	1.120913737	
P(T<=t) one-tail	0.139433694	
t Critical one-tail	1.745883676	
P(T<=t) two-tail	0.278867387	
t Critical two-tail	2.119905299	

this is "by-hand"
What can we conclude?
The p-value is higher than 0.05. There is not sufficient evidence in this dataset to conclude that men and women are paid differently.

For doing confidence intervals on this same type of data, you can them… use the same standard error formulas for those confidence intervals, with the mean (center) of the interval being the difference between the two samples, and the same methods for finding the t-score or z-score to multiply by the standard error.

Hypothesis tests for multiple groups (ANOVA)
Goodness-of-fit tests
Tests of Independence

Regression