proprep™

Workbook

Table of Contents

Hypothesis Testing about Difference between Means

Known Population Variances

Theory

What are we doing?
- We are trying to compare two population means:

$$H_0 \quad \mu_1 - \mu_2 = c$$
$$H_1 \quad \mu_1 - \mu_2 \neq c$$

What are our conditions?

1) Independent samples.
2) $\sigma_1, \ \sigma_2$ are known.
3) Variables follow normal distribution, or we have sample sizes of at least 30.

What statistic do we use?

$$Z_{\bar{x}_1 - \bar{x}_2} = \frac{(\bar{x}_1 - \bar{x}_2) - c}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Example (Solution in the video)

Suppose that in 2004, men, on average, earned $\$750$ more in weekly salary.
Let's say that the standard deviations of weekly earnings are:
$\$750$ for men and $\$500$ for women.
We want to check if the salary gap has narrowed since 2004.
We sample 100 men and find they earn on average $\$2268$.
We sample 80 women and find they earn on average $\$1953$.
What conclusion can be drawn, at a 5% level of significance?

Questions

1) A study asserts that people on the East Coast watch more TV, on average, than people in the Midwest.
100 people from the East Coast and 107 from the Midwest were sampled:
The average TV watching time from the East Coast sample was 2.7 hours.
The average TV watching time from the Midwest sample was 1.8 hours.
Assuming a standard deviation of 1 hour, test the study's assertion at a 1% level of significance.

2) A certain test has scores distributed normally with a standard deviation of 100.
A test prep company claims that it can improve the average score by at least 30 points.
A sample of 25 people who worked with the test prep company had an average score of 561.
A sample of 20 people who did not work with the test prep company had an average score of 508.
At a 5% level of significance, what conclusion can be drawn?

3) The daily output of a factory on the day shift was recorded for a random sample of 20 days, and the resulting average was 340 products a day.
The daily output of the factory on the night shift for a random sample of 20 other days, and the resulting average was 295 products a day. Assume that the standard deviation of daily output is 40 products during the day and 30 products at night.
At an 8% level of significance, what conclusion can be drawn?

4) A study in Europe established that men are 8 cm taller than women on average.
We wish to investigate whether the gender gap in height is bigger in the USA.
40 men and 40 women were randomly sampled.
Assume that the standard deviation of heights is 6 cm among women and 12 cm among men.

a. What are the study hypotheses, and what is the decision criterion at a 10% level of significance?
b. If the height gap between men and women in the USA is actually 11 cm, what is the probability that the study will not detect this bigger height gap?
What is this probability called?

Answer Key

1) We will reject H_0.

2) We will not reject H_0.

3) We will reject H_0.

4) a. We will reject H_0. b. If the average height of the men in the sample is more than 10.72 cm higher than the average height of the women in the sample.
0.4483 is called the Probability of a Type II error.

Unknown Population Variances

Example (solution in the video)

Is there a difference (at 5% significance) between the average psychometric test scores of college and high school students?

	College	H.S.
Sample size	46	50
Average	543	508
Standard Deviation	123	178

Questions

1) A construction engineering company wants to compare the strength of two bolts. The average strength and standard deviation in a sample of 38 type A screws were 28 units and 4 units, respectively. The average and standard deviation in a sample of 45 type B screws were 25 units and 6 units, respectively.
Based on the sample results, is there a difference between the two types of bolts? Test at a 5% level of significance.

2) In order to check whether drivers driving under the influence of alcohol drive faster, each driver's maximum speed was recorded. The following table displays the results. At a 5% level of significance, what conclusion can be drawn?

	Sample Size	$\overline{X}$	S
Drivers under the influence ($\mathrm{km/hr}$)	70	80	20
Drivers not under the influence ($\mathrm{km/hr}$)	100	60	15

Answer Key

1) We will reject H_0.

2) a. 0 b. We will reject H_0.

Unknown Population Variances assumed to be Equal

We are trying to estimate the difference between two population means.
If our samples are not large and our variable follows a normal distribution, we use a two-sample t-test.

If we can assume that the population variances are equal,
we "pool" the variances from the samples.

Null Hypothesis:	$H_0 \quad \mu_1 - \mu_2 = c$	$H_0 \quad \mu_1 - \mu_2 = c$	$H_0 \quad \mu_1 - \mu_2 = c$
Alternative Hypothesis:	$H_1 \quad \mu_1 - \mu_2 \neq c$	$H_1 \quad \mu_1 - \mu_2 < c$	$H_1 \quad \mu_1 - \mu_2 > c$

$$t_{\bar{x}_1 - \bar{x}_2} = \frac{(\bar{x}_1 - \bar{x}_2) - c}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Example (solution in the video)
A company claims that it has a metal whose melting temperatures higher than the melting temperature of the current one.
Assume that melting temperatures are distributed normally with equal variances.
Test the company's claim at a 5% level of significance.

	Old type	New type
Sample size	10	12
Average melting temp	1,170 degrees	1317 degrees
Variance of melting temp	200	260

Questions

1) The following table displays the figures for areas of apartments built in 2012 and 2013: (in square meters)

2012	120	112	95	130	90	94	120
2013	100	82	91	105	74	69	

Test at a 5% level of significance whether there was a significant drop in 2013 in the area of apartments, compared with 2012.
Assume that the variable of apartment area is distributed normally and that the variance of the areas in the two years are equal.

2) The following table displays the results of a sample of the lifespan of $W60$ and $W100$ light bulbs, in hours
Assume that the lifespan is distributed normally and equal variances in the lifespan of the two types.

a. Test at a 5% level of significance whether light bulbs last longer on the average than $W100$ light bulbs.
b. At a 5% level of significance, test whether $W60$ light-bulbs last longer than 1,000 hours.

Group	$2-60W$	$1-100W$
$\bar{x}$	1007	956
S	80	72
n	13	15

Answer Key

1) We will reject H_0.

2) a. We will reject H_0. b. We will not reject H_0.