study unit fifteen

hypothesis tests applied to means: two independent samples

In study unit 14 we did hypothesis testing with two related samples. Now we are going to talk about hypothesis tests when dealing with two independent samples. We shall go through the same steps of hypothesis testing, except that we shall be using a different type of test statistic. We shall also look at the adjustments we make when there are differences in homogeneity of variance and when samples are not the same size.

What is the purpose of applying hypothesis tests to two independent samples?

The purpose of hypothesis testing with two independent means is to help you decide whether an observed difference between two sample means is accidental or whether it represents a real difference between populations. In hypothesis testing terms, the purpose of the *t*-test is to help us decide whether or not to reject a null hypothesis postulating no difference between the means.

What are two independent samples?

activity

Do you still remember when we classify samples as related? The exact opposite applies to independent samples.

First make a list of the characteristics of related samples. Then complete the table for independent samples. (Read Howell's introduction on p 259 to help you with this activity.)

Related samples	Independent samples

Independent samples are samples that are not related in any way, or when there is no correlation between individual pairs of participants in two samples — that is, when participants are included in only one of the two samples in an experiment.

activity

To help you understand the principle underlying hypothesis testing when dealing with two independent samples you should work through the introductory part to chapter 14 in Howell. Follow the reasoning about sampling distributions of differences between means. (Do you remember what a sampling distribution is? If not, refresh your memory by looking at study unit 13. As you will see when you read the introduction to Howell's chapter 14 and the introductory part of study unit 13, we are now dealing with the sampling distribution of the difference between means.)

1 Complete the following sentence:
The sampling distribution of the difference between means is

____

____

		True	False
2	Because we sample each population independently, the sample means will also be independent.		
3	The mean of the sampling distribution will be $\mu_1 - \mu_2$.		

1 You will find the answer on page 260 in Howell.
2 True
3 True

How do we compute the *t*-test for independent samples?

Now that you know what independent samples are and why we apply a *t*-test to them, the next step is to determine whether there is a significant difference between the means of two independent samples. Again, this is quite a simple computation when you use the formula that Howell provides above "Pooling Variances" page 263. You don't have to know the background; merely learn the formula.

activity

Work through the example which follows and make sure how this test statistic (in this case the *t*-test for independent groups) is computed.

Prof Martins wants to determine whether there is a significant difference between men's and women's performance scores question 5 in the IOP201-Q exam. She gets the scores of 20 randomly selected students and decides to test significance at the 5% and 1% levels.

Data

Men		**Women**	
X	X^2	Y	Y^2
4	16	2	4
3	9	1	1
2	4	5	25
5	25	3	9
1	1	3	9
6	36	4	16
2	4	4	16
1	1	1	1
7	49	5	25
6	36	5	25
$\Sigma X = 37$		$\Sigma Y = 33$	
$\Sigma X^2 = 181$		$\Sigma Y^2 = 131$	
$\bar{X}$ 3,7		3,3	
s^2 4,88		2,47	
N 10		10	

The formula which we use to compute the *t*-value for two independent groups requires us to use the variance of both groups. If you are not given the variance, you must first compute it before you can use the *t*-test. Work through the example to see how this is done.

$$s_X^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N - 1}$$
$$= \frac{181 - \frac{(37)^2}{10}}{9}$$
$$= \frac{181 - \frac{1369}{10}}{9}$$
$$= \frac{181 - 136,9}{9}$$
$$= \frac{44,1}{9}$$
$$= 4,88$$

$$s_Y^2 = \frac{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}}{N - 1}$$
$$= \frac{131 - \frac{(33)^2}{10}}{9}$$
$$= \frac{131 - \frac{1089}{10}}{9}$$
$$= \frac{131 - 108,9}{9}$$
$$= \frac{22,1}{9}$$
$$= 2,46$$

Now that you have the values for both variances, you can compute the *t*-value.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$
$$= \frac{3,7 - 3,3}{\sqrt{\frac{4,88}{10} + \frac{2,46}{10}}}$$
$$= \frac{0,4}{\sqrt{0,488 + 0,246}}$$
$$= \frac{0,4}{\sqrt{0,734}}$$
$$= \frac{0,4}{0,857}$$
$$= 0,47$$

$$df = N_1 + N_2 - 2$$
$$= 10 + 10 - 2$$
$$= 18$$

$t_{0.05}$ (18) = 2,101

From Table E.6

$t_{0.01}$ (18) = 2,878

For α = 0,05	For α = 0,01
0,47 < 2,101	0,47 < 2,878
∴ Do not reject H_0	∴ Do not reject H_0

Interpretation

Prof Martins concludes with 99% certainty that there is no difference between the achievement of men and woman on Question 5 of the examination paper.

This is the most basic variation of the *t*-test for independent groups: one in which there are no differences between the number (*N*) of participants and also no large difference in variances. In the next example we are going to look at these aspects and see how we then do the computation slightly differently.

Why do we pool variances?

When we are dealing with samples of unequal size, we have to pool the variances of the two samples. Howell explains this on pages 263 and 264 and also gives the background to homogeneity of variances, on which the formulas are based. Make sure that you can use the formula for pooled variances by thoroughly studying Howells's example.

How do we determine the degrees of freedom?

The degrees of freedom for the *t*-test for independent samples are computed differently from those for related samples. The formula is $N_1 + N_2 - 2df$. Make sure that you can compute it — you will find it in the paragraph "Degrees of freedom for *t*" on page 264 in Howell.

activity

To familiarise yourself with the various facets of the use of the *t*-test, it will help you to go through the example which Howell provides in section 14.1 (p 265–267). Follow the steps of hypothesis testing, which you should know quite well by now. Howell's presentation is again unstructured, but he does spell out the various steps. Then do the following exercises:

		True	False
1	Howell uses a one-tailed test.		
2	The degrees of freedom of the two groups are not summed.		
3	There is no difference between the means of the populations.		

Now that you have gone through a complete example, try to do the example in section 14.4 of Howell by yourself.

- Follow all the steps in the process, including the formulation of null and alternative hypotheses and the decision on a one- or two-tailed test.
- Use your pocket calculator to complete the steps in the computation of the test statistic and determine the degrees of freedom.
- Look up the critical values in the table and apply the decision-making rules.
- Write down your interpretation concerning rejection/nonrejection of the null hypothesis.

---

1 False
2 False
3 False

What happens if the variances of the two samples differ greatly?

Sometimes the variance of one sample may be up to four times the variance of the other sample. In such cases we have to compensate for this.

activity

Howell explains heterogeneity of variance in section 14.2. Make sure that you know how to compensate for it by adjusting the degrees of freedom.

Now that you have worked through this and are familiar with the concept of heterogeneity of variance, study the rest of section 14.2 in Howell, then complete the following sentences:

1 Heterogeneity of variance is

____

____

2 I would compensate for it by

____

____

3 Work through the example in 14.4 and look up Howell's answers.

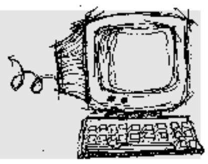
---

1 You will find the answer on page 267 in Howell.

2 We compensate for heterogeneity by computing the t-test separately for the different variances (not pooling them). The smaller N is then used to read the degrees of freedom. Remember that the computer can compensate for heterogeneity, but this is the only way of doing it manually.

You have learnt quite a few new skills and at this stage you should be able to

- know *when* to use an independent t-test
- test a hypothesis by *applying* the t-test to independent samples and *compute* a test statistic, even when sample sizes are not the same or when the variance of one sample is four times that of the other sample
- come to a final *conclusion* about your research hypothesis



Go through section 14.6 in Howell and make sure that you can identify the t-value on the printouts.

activity

Now that you have gone through all the alternatives and know when to pool variances and how to compensate for heterogeneity of variance, try to do the following exercises in Howell's section 14.9:

1 14.1

2 14.7

---

1 See *Student handbook*.

2 $t = 2,66$. Reject the null hypothesis. See *Student handbook*.

Nonparametric equivalent of the *t*-test for independent samples

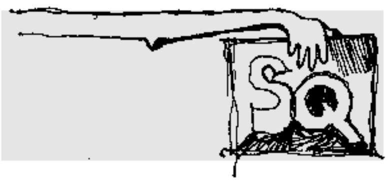
In this section we deal theoretically with the nonparametric equivalent of the t-test for two independent samples by looking at the Mann-Whitney test. As an industrial psychologist you have to know what to do when your data are not normally distributed. That is what we explain in this section.

activity

There is not a great deal to learn in this section. Study the first paragraph of section 20.1 in Howell.

Go back to study unit 14 and enter the new information in the table at the end of that unit. In this way you will enter all the possible nonparametric equivalents and be able to use the table as a quick, complete source of reference.

Now that you have completed this section you should be able to *name* the nonparametric equivalent of the t-test for two independent samples. Remember, you need not be able to describe it or know the formula.



Exercises 14.10, 14.13 and 14.14

(Exercise 14.14: Note that you have to recode the ADDSC scores.)

The output for exercise 14.10 looks like this:

Variable	Obs	Mean	Std. Dev.
Good	9	18.77778	3.929942
Poor	8	17.625	4.172615
combined	17	18.23529	3.961209

Note the descriptive statistics also given for each group

```
Ho:  mean(x) - mean(y)    = 0  (assuming equal variances)
                     t    = 0.59 with 15 d.f.
              Pr > |t|    = 0.5663
    95% conf. interval    = (-3.0366, 5.3422)
```

Do not reject *Ho*

Exercise 14.13: t = 1,66 (See *Student handbook)*.

Exercise 14.14: t = 3,76