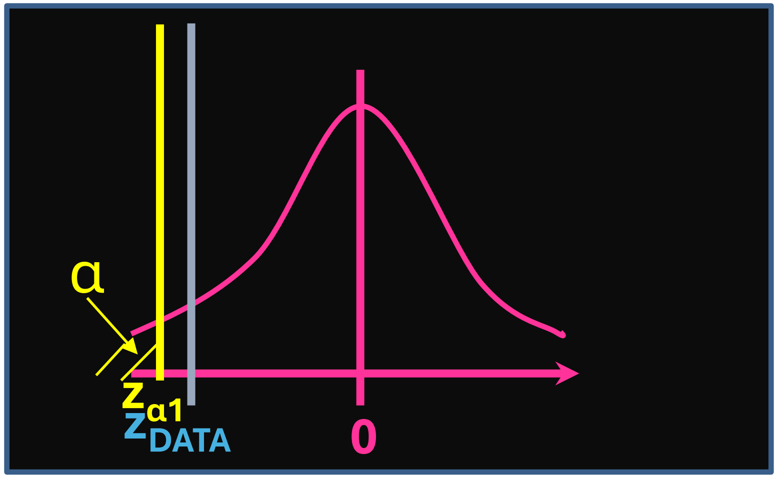
More hypothesis testing
p-values

Lecture 15

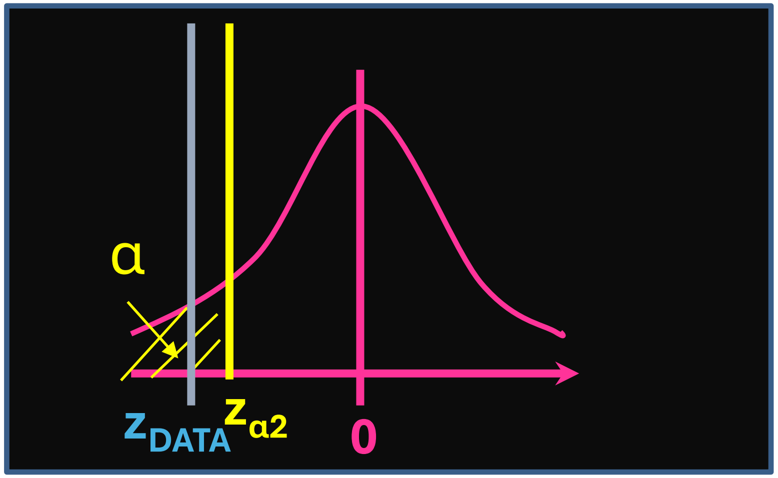
Different results for different α

- Sometimes we are not sure what value of significance to use

Example: left-tailed test



Failed to reject Ho at significance level α1



Reject Ho at a different significance level α2

For the same sample

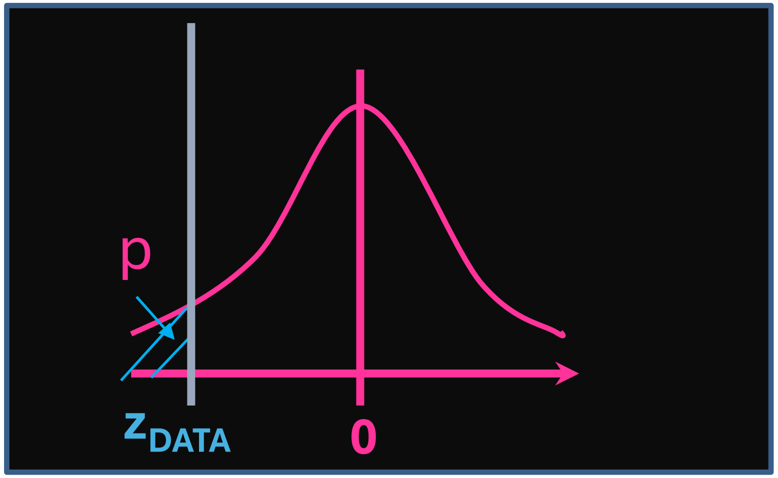
Hypothesis testing using p-values

- *p-value* - probability of obtaining a sample more extreme that the current data assuming that Ho is true
- How likely your data and beyond is observed under a null hypothesis

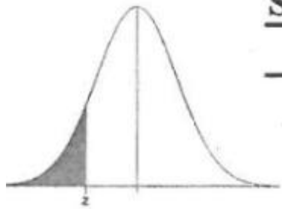
Hypothesis testing using p-values

- *p-value* - probability of obtaining a sample more extreme that the current data assuming that Ho is true
- How likely your data and beyond is observed under a null hypothesis

Example: left-tailed test



p-value lets you specify the level of significance **from data**, instead of testing current data with different α-thresholds



Finding probability of observing sample $\leq z_{DATA} = -1.34$

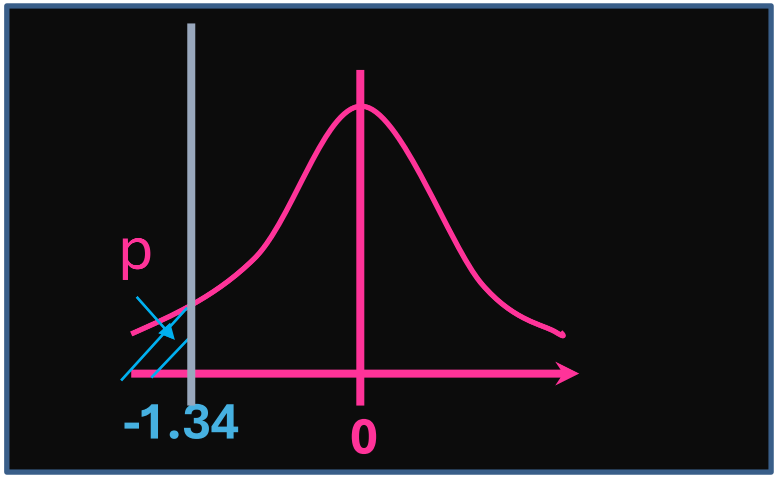
STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to t

Z	.00	.01	.02	.03	.04	.05	.06
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383

Ex. 7: hypothesis testing with p-value

- Ho: $\mu \geq 0.15$
- Ha: $\mu < 0.15$

$z_{DATA} = -1.34$



- $p(z<-1.34) = 0.00901$ (see prev. slide)

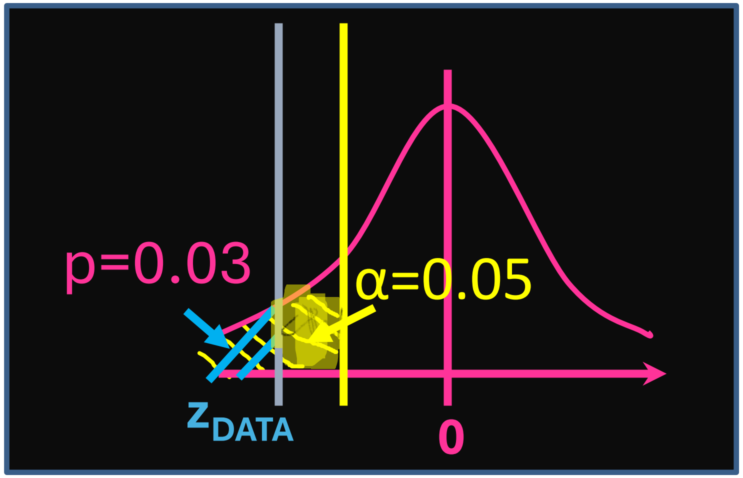
- This means that if Ho is true then probability of observing this sample is <=0.009
- This is a very low probability, and we should probably reject Ho

Ex. 8:
hypothesis testing with p-value and alpha

If $p \leq \alpha$ => reject Ho
If $p > \alpha$ => failed to reject Ho

- Ho: $\mu \geq 409$ $z_{DATA} = -1.87$
- Ha: $\mu < 409$ $\alpha = 0.05$

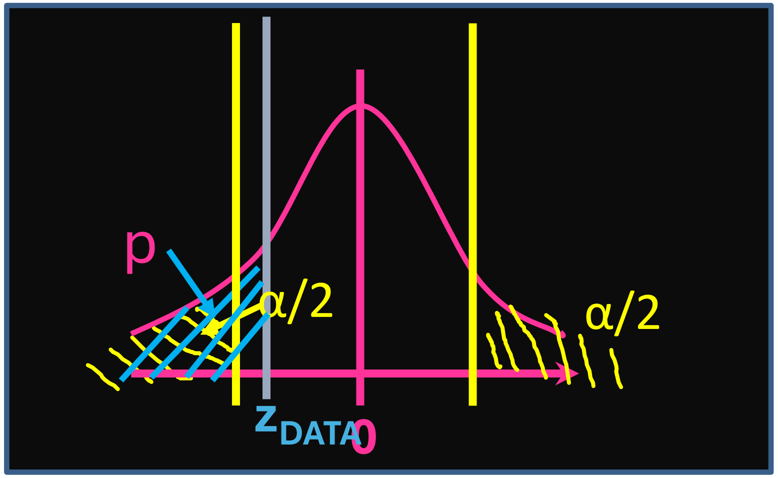
- $p(z<-1.87) = 0.0307$ (from z-table)



Reject Ho

Ex. 8: p-value for a two-tailed test

If p ≤ α/2 => reject Ho
If p > α/2 => failed to reject Ho



Failed to reject Ho

- Ho: μ = 2
- Ha: μ ≠ 2

$z_{DATA} = -2.28$

$\alpha = 0.02$

p(z<-2.28) = 0.013 (from z-table)

Finding $t_{\alpha,DOF}$ from t-table

df	0.1	0.05	0.025	0.02	0.01	0.005
1	3.078	6.314	12.706	15.895	31.821	63.657
2	1.886	2.920	4.303	4.849	6.965	9.925
3	1.638	2.353	3.182	3.482	4.541	5.841
4	1.533	2.132	2.776	2.999	3.747	4.604
5	1.476	2.015	2.571	2.757	3.365	4.032
6	1.440	1.943	2.447	2.612	3.143	3.707
7	1.415	1.895	2.365	2.517	2.998	3.499
8	1.397	1.860	2.306	2.449	2.896	3.355
9	1.383	1.833	2.262	2.398	2.821	3.250
10	1.372	1.812	2.228	2.359	2.764	3.169
11	1.363	1.796	2.201	2.328	2.718	3.106
12	1.356	1.782	2.179	2.303	2.681	3.055
13	1.350	1.771	2.160	2.282	2.650	3.012
14	1.345	1.761	2.145	2.264	2.624	2.977
15	1.341	1.753	2.131	2.249	2.602	2.947
16	1.337	1.746	2.120	2.235	2.583	2.921
17	1.333	1.740	2.110	2.224	2.567	2.898
18	1.330	1.734	2.101	2.214	2.552	2.878
19	1.328	1.729	2.093	2.205	2.539	2.861
20	1.325	1.725	2.086	2.197	2.528	2.845
21	1.323	1.721	2.080	2.189	2.518	2.831
22	1.321	1.717	2.074	2.183	2.508	2.819
23	1.319	1.714	2.069	2.177	2.500	2.807
24	1.318	1.711	2.064	2.172	2.492	2.797
25	1.316	1.708	2.060	2.167	2.485	2.787
26	1.315	1.706	2.056	2.162	2.479	2.779
27	1.314	1.703	2.052	2.158	2.473	2.771
28	1.313	1.701	2.048	2.154	2.467	2.763

n=29, DOF = 28
α = 0.01

Ex. 9: mean, small samples, rejection regions

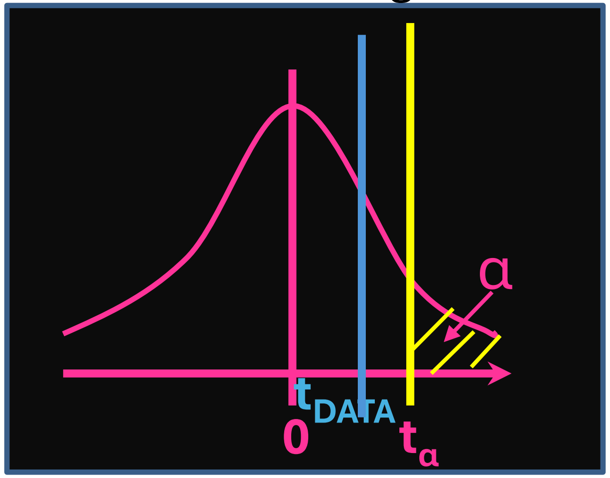
Right-tailed test

- Ho: $\mu \leq 28$ n=29 $t_{DATA} = 1.99$
- Ha: $\mu > 28$ $\alpha = 0.01$

- DOF = n-1 = 28
- $t_\alpha = t_{0.01\ @\ 28\ DOF} = 2.47$ (see prev. slide)

- Failed to reject Ho

T-table always gives the t_α for α to the right



Method of rejection regions

Finding p-value from t-table

df	0.1	0.05	0.025	0.02	0.01	0.005
1	3.078	6.314	12.706	15.895	31.821	63.657
2	1.886	2.920	4.303	4.849	6.965	9.925
3	1.638	2.353	3.182	3.482	4.541	5.841
4	1.533	2.132	2.776	2.999	3.747	4.604
5	1.476	2.015	2.571	2.757	3.365	4.032
6	1.440	1.943	2.447	2.612	3.143	3.707
7	1.415	1.895	2.365	2.517	2.998	3.499
8	1.397	1.860	2.306	2.449	2.896	3.355
9	1.383	1.833	2.262	2.398	2.821	3.250
10	1.372	1.812	2.228	2.359	2.764	3.169
11	1.363	1.796	2.201	2.328	2.718	3.106
12	1.356	1.782	2.179	2.303	2.681	3.055
13	1.350	1.771	2.160	2.282	2.650	3.012
14	1.345	1.761	2.145	2.264	2.624	2.977
15	1.341	1.753	2.131	2.249	2.602	2.947
16	1.337	1.746	2.120	2.235	2.583	2.921
17	1.333	1.740	2.110	2.224	2.567	2.898
18	1.330	1.734	2.101	2.214	2.552	2.878
19	1.328	1.729	2.093	2.205	2.539	2.861
20	1.325	1.725	2.086	2.197	2.528	2.845
21	1.323	1.721	2.080	2.189	2.518	2.831
22	1.321	1.717	2.074	2.183	2.508	2.819
23	1.319	1.714	2.069	2.177	2.500	2.807
24	1.318	1.711	2.064	2.172	2.492	2.797
25	1.316	1.708	2.060	2.167	2.485	2.787
26	1.315	1.706	2.056	2.162	2.479	2.779
27	1.314	1.703	2.052	2.158	2.473	2.771
28	1.313	1.701	2.048	2.154	2.467	2.763

n=29

$t_{DATA} = 1.99$

Ex. 9A: mean, small samples, p-value

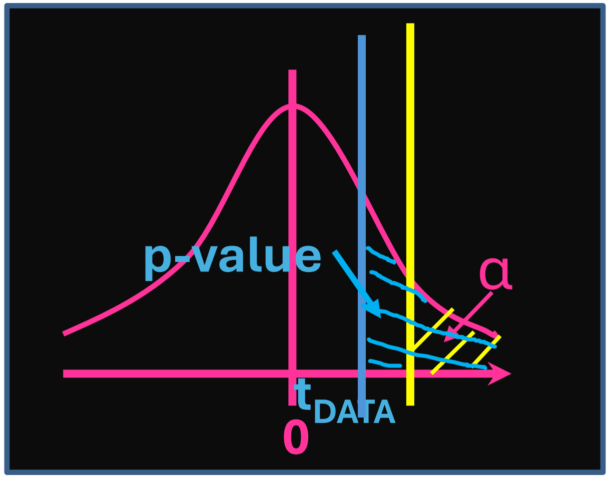
Right-tailed test

- Ho: $\mu \leq 28$ n=29 $t_{DATA} = 1.99$
- Ha: $\mu > 28$ $\alpha = 0.01$

- DOF = n-1 = 28
- p-value ($t_{1.99,\ DOF\ =\ 28}$) ≈ 0.03 (see prev. slide)

- p-value > α

- Failed to reject Ho

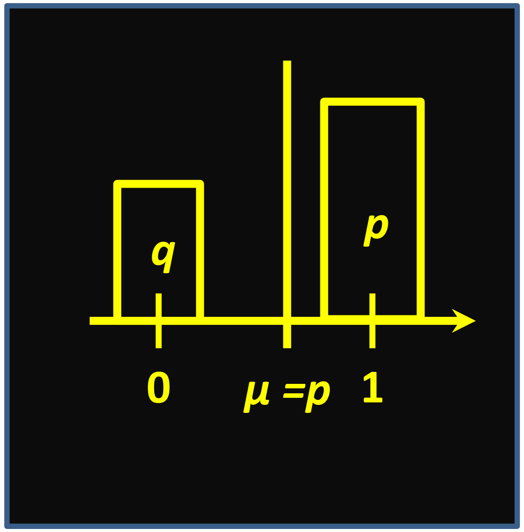


Using p-value

Population proportions

Population proportions

- Population proportion is a percentage of Bernoulli experiments with an outcome "yes"(success) - a percentage of a population behaving a certain way
- The proportion (the probability of success) is denoted by small letter *p*. The probability of failure by letter $q = 1-p$
- The expected value (mean) of this distribution: $\mu = p$
- The variance $\sigma^2 = pq$,
- The standard deviation $\sigma = \sqrt{p(p-1)}$



Hypotheses with population proportions

- Proportion of a sample is denoted as: $\hat{p}$

- We can approximate the distribution of proportions (Bernoulli means) using normal distribution when:
  np ≥ 5
  n(1-p) ≥ 5

Then the hypothesis test statistic is:

$$z_{DATA} = \frac{\overset{\bar{x}}{\hat{p}} - \overset{\mu}{p}}{\sqrt{\frac{\overset{\sigma^2}{\hat{p}(1-\hat{p})}}{n}}}$$

Compare to:

$$z_{DATA} = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}}$$

For population proportion

For population mean

$$z_{DATA} = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Ex. 10: proportions, p-value

- A report states that at least 75% of women like red roses.
- Alice claims this figure is too high.
- She asks 125 women and finds that 92 do like red roses.
- At a 0.10 level of significance test the claim.

$$z_{DATA} = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Ex. 10. proportions, p-value

- A report states that at least 75% of women like red roses.
- Alice claims this figure is too high. She asks 125 women and finds that 92 do like red roses.
- At a 0.10 level of significance test the claim.

Ho: p >= 0.75
Ha: p < 0.75
Left-tailed test

From a sample:
$\hat{p} = \frac{92}{125} = 0.736$,
s^2= 0.736*(1-0.736) = 0.19
Can we use z-distribution?
np = 92 ≥ 5
n(1-p) = 125*0.264 = 33 ≥ 5 Yes!
z_{DATA} = -0.362
p-value = p(z<-0.362) = 0.3594 > α = 0.1

Failed to reject Ho

Comparing two populations

Comparing means of two populations: large independent samples

- If n >=30 => use normal distribution

We want to compare:

- μ_1 – population mean 1
- μ_2 – population mean 2

Hypotheses of type:

- Ho: $\mu_1 \le \mu_2$
- Ha: $\mu_1 > \mu_2$

Usually written as:

- Ho: $\mu_1 - \mu_2 \le 0$
- Ha: $\mu_1 - \mu_2 > 0$

Then the test-statistic is:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Ex. 11

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

A company believes that employees who drink coffee **do not complete** more tasks per day than those who don't drink coffee.

- Sample 1 (Non-coffee drinkers): n_1 = 40, $\bar{x}_1$ = 16.9 tasks, s_1 = 3.2
- Sample 2 (Coffee drinkers): n_2 = 45, $\bar{x}_2$ = 18.2 tasks, s_2 = 3.5

At α = 0.05, test if coffee drinkers are more productive.

Ex. 11

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

A company believes that employees who drink coffee **do not complete** more tasks per day than those who don't drink coffee.

- Sample 1 (Non-coffee drinkers): $n_1 = 40$, $\bar{x}_1 = 16.9$ tasks, $s_1 = 3.2$
- Sample 2 (Coffee drinkers): $n_2 = 45$, $\bar{x}_2 = 18.2$ tasks, $s_2 = 3.5$

At $\alpha = 0.05$, test if coffee drinkers are more productive.

- Ho: $\mu_1 \geq \mu_2 \Leftrightarrow \mu_1 - \mu_2 \geq 0$ (no difference in drinkers-non-drinkers)
- Ha: $\mu_1 < \mu_2 \Leftrightarrow \mu_1 - \mu_2 < 0$ (left-tailed test)

- Since under Ho the difference between μ_1 and μ_2 should be zero, the formula is simplified and we compute:
- $z_{DATA} = -1.79$
- $z_C = -1.645$
- Since $z_{DATA} < z_C$ we **reject H_0** — coffee drinkers are indeed more productive.

Comparing **two** proportions: large independent samples

- Independent (non-overlapping) samples
- For both samples:
  - n$\hat{p}$ >= 5
  - n(1- $\hat{p}$) >= 5

- Test statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}(1-\bar{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

where: $\bar{p} = \dfrac{\overset{p_1}{x_1} + \overset{p_2}{x_2}}{n_1 + n_2}$

Ex. 12: Vaccine Effectiveness

In a study, 400 people receive a new vaccine and 400 receive a placebo.

- Vaccine group: 18 people get sick ($p1=18/400$)
- Placebo group: 45 people get sick ($p2=45/400$)

At $\alpha = 0.05$, test whether the vaccine reduces the infection rate.

Ex. 12: Vaccine Effectiveness

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}(1-\bar{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

In a study, 400 people receive a new vaccine and 400 receive a placebo.

- Vaccine group: 18 people get sick (p1=18/400)
- Placebo group: 45 people get sick (p2=45/400)

At α = 0.05, test whether the vaccine reduces the infection rate.

- Ho: $p_1 \geq p_2 \Leftrightarrow p_1 - p_2 \geq 0$
- Ha: $p_1 < p_2 \Leftrightarrow p_1 - p_2 < 0$

- For a **left-tailed** test at α = 0.05: $z_c = -1.645$
- $z_{DATA} = -3.55 < -1.645$

- **Reject H_0**
- There is strong evidence (p ≈ 0.0002) that the **vaccine significantly reduces infection rates** compared to placebo.

Dependent samples: paired data

- In many studies we have the statistics about **the same** group of people: clinical studies, before/after
- Hypothesis testing with **paired data**

Ex. 13: paired data, small samples

weight loss drug

		Person 1	2	3	4	5
Population 1	Before	160	175	182	180	160
Population 2	After	155	158	179	179	159
	d	-5	-17	-3	-1	-1

- What is important here is the difference, not absolute values:

d = after – before

- We study mean of paired differences: $\bar{d} = \frac{\sum d}{n}$

- The means of differences follow t-distribution

- The test statistic is : $$t = \frac{\bar{d} - \mu}{\frac{s_d}{\sqrt{n}}}$$ DOF = n-1 Use rejection regions

Ex. 13: solution

- We test whether the drug reduces weight on average.
- $\mathrm{Ho}: \mu_d \geq 0$ *(no decrease in mean weight)*
- $\mathrm{Ha}: \mu_d < 0$ *(weight decreases after treatment)*

where μ_d is the population mean of the paired differences (After – Before).

- Compute sample mean: $\bar{x}_{\mathrm{d}} = -5.4$
- Compute sample standard deviation: $s_d = 6.69$
- $t_{DATA} = (-5.4 - 0)/(6.69/\mathrm{sqrt}(5)) = -1.81$
- $t_{\alpha=0.05\ @\ DOF=4} = -2.132$
- $t_{DATA} = -1.81 > -2.132$ => We **failed to reject** Ho

There isn't enough evidence to conclude that the drug causes significant weight loss (at the 5% level of significance).