3. **Test Statistic:**

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \quad \text{where } \hat{p} = \frac{84+132}{100+150} = \frac{216}{250}$$

$$= \frac{\frac{84}{100} - \frac{132}{150}}{\sqrt{\frac{216}{250}\left(\frac{34}{250}\right)\left(\frac{1}{100}+\frac{1}{150}\right)}}$$

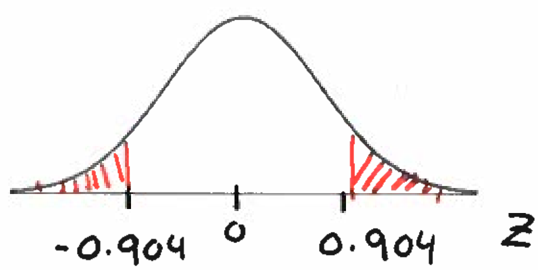
$$= \frac{-0.04}{0.044254} = -0.904$$

4. ***P*-value:**

$$2P(z < -0.904)$$

$= 2(0.1841)$ from z-table

$= 0.3682 > 0.05$



$p_1 - p_2$: Difference in proportions
$H_0 : p_1 - p_2 = 0$
$H_A : p_1 - p_2 \neq 0$

$z_0^2 = (-0.90387691)^2 = 0.81699346 = \chi_0^2$

Hypothesis test results:

Difference	Count1	Total1	Count2	Total2	Sample Diff.	Std. Err.	Z-Stat	P-value
$p_1 - p_2$	84	100	132	150	-0.04	0.044253813	-0.90387691	0.3661

Same p-value as χ^2-test

5. **Conclusion:** Since P-value > 0.05, we do not reject H_0 at the 0.05 significance level, that is, there is not enough statistical evidence to conclude that the rate of germination is different between the treated and untreated seeds.

Example: Identify which chi-squared test should be used in each of the following situations:

a) Medical researchers followed 6272 men for 30 years to see if there was any association between fish consumption (never consume, small part of diet, moderate part of diet, large part of diet) and incidence of cancer.

1 population
2 variables

Independence

b) Researchers gathered data on employment status of adults in five Scandinavian countries by taking random samples of adults from each country. They want to know if there is sufficient evidence to conclude that a difference exists in the unemployment rates of the five countries.

5 groups/populations
1 variable

Homogeneity

c) A major roadway with four lanes in each direction was studied to determine if drivers prefer to drive on the inside lanes. A total of 1000 vehicles were observed during morning traffic and the number of cars in each lane was recorded: 294 in the inner most lane, 276 in second lane, 238 in the third lane, and 192 in the outer most lane. Do the data present sufficient evidence to indicate that some lanes are preferred over others?

1 population
1 variable with
4 categories

Goodness-of-fit:

$H_0: p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$

Chapter 17: More About Tests

When performing a hypothesis test, there are two types of errors that can occur:

1. **Type I Error**: the null hypothesis is true, but we reject it.
2. **Type II Error**: the null hypothesis is false, but we fail to reject it.

The Truth

		H_0 True	H_0 False
Our Decision	Reject H_0	Type I Error	Correct
	Fail to Reject H_0	Correct	Type II Error

Example: In a court trial, if

H_0: person is innocent
H_A: person is guilty

a) What is a Type I error in this context?

H_0 true = innocent
reject H_0 = found guilty

convicted an innocent person

b) What is a Type II error in this contest?

H_0 false = person not innocent = guilty
not reject H_0 = found not guilty

let guilty person go free.

c) Which type of error is more serious?

??

perhaps the question is best left to philosophers

Example: A calculator manufacturer receives very large shipments of circuits from a supplier. It is too costly and time consuming to inspect all circuits in a shipment, so they take a sample of each shipment to inspect. Data is obtained from the sample to test the hypotheses:

$$H_0 : p \overset{\le}{=} 0.05$$
$$H_A : p > 0.05$$

where p is the proportion of defectives in the shipment. If H_0 is rejected, the shipment is sent back to the supplier. If H_0 is not rejected, the circuits are used in the production of calculators.

a) What type of error occurred if the shipment is returned, but contains only 4% defective circuits?

returned = H_0 rejected   Type I

4% defective : $0.04 \not> 0.05$

b) What type of error occurred if the shipment contains 6% defective circuits, but is used to make calculators?

Shipment retained = H_0 not rejected   Type II

6% defective = H_0 false

c) From the manufacturer's point of view, which type of error would be more serious?

Type II

d) From the supplier's point of view, which type of error would be more serious?

Type I

Example: Water samples are taken from water that is used for cooling as it is returned to a river from a power plant. As long as the mean temperature of the discharged water is at most 65 degrees Celsius, no harm will be done to the river's ecosystem. Fifty samples will be taken and measured at randomly selected times and the resulting data will be used to test the hypotheses:

$$H_0 : \mu = 65 \text{ degree Celsius}$$
$$H_A : \mu > 65 \text{ degrees Celsius}$$

a) What is a Type I error in this context?

concluding water discharged is too hot when mean temperature is okay.

b) What is a Type II error in this context?

Concluding mean temperature is okay when it is actually too hot.

c) Which type of error is more serious?

Type II → river's ecosystem harmed.

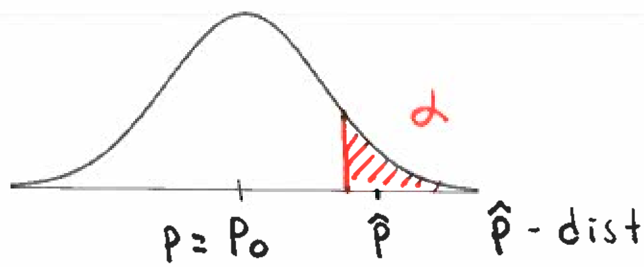
Probability of Errors

The **probability that a Type I error** occurs is the significance level α, that is

$$\alpha = P(\text{Type I Error}) = P(H_0 \text{ is rejected} | H_0 \text{ is true})$$

ex: $H_0: p = p_0$

$H_A: p > p_0$



The **probability that a Type II error** occurs is denoted β, that is,

$$\beta = P(\text{Type II Error}) = P(H_0 \text{ is not rejected} | H_0 \text{ is false})$$

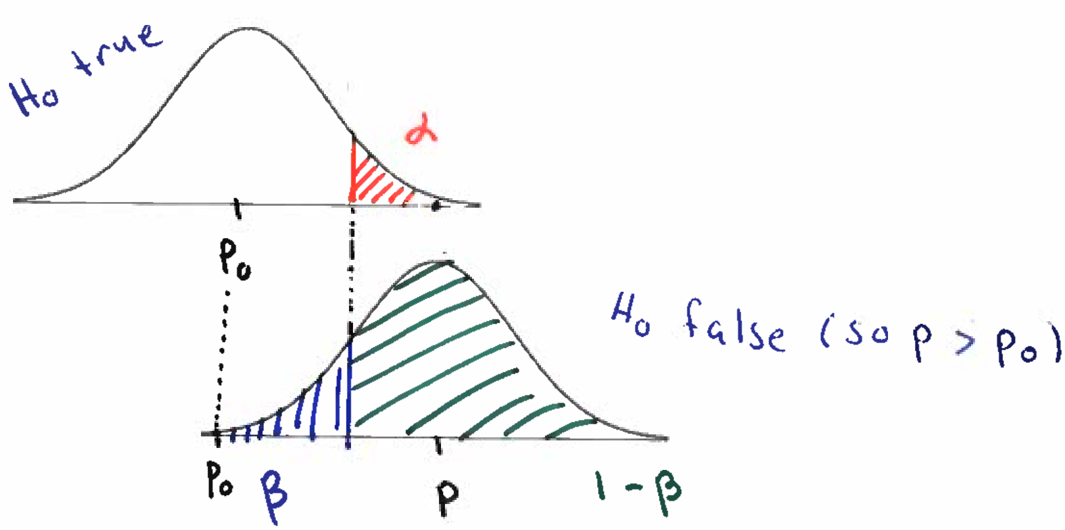
The value $1 - \beta$ is the **power** of the test. It is the probability of rejecting H_0 when H_0 is false, that is,

ex.

$H_0: p = p_0$

$H_0: p > p_0$

$$1 - \beta = P(H_0 \text{ is rejected} | H_0 \text{ is false})$$



$\beta + \text{power} = 1$

$\alpha \uparrow \quad \beta \downarrow \quad$ power $\uparrow$

$\alpha \downarrow \quad \beta \uparrow \quad$ power $\downarrow$

Note:

- As α increases, (n fixed)
  - the probability of a Type I error increases
  - the probability β of a Type II error decrease
  - the power of the test increases
- The value of α is chosen by the researcher.
- The value of β is hard or even impossible to compute, since we do not know the actual value of the parameter.

To minimize both α, β
- choose α small
- make n large.

- As the sample size increases
  - the probability of a Type I error stays the same.
  - the probability β of a Type II error decreases.
  - the power of the test increases.
- The **effect size** is the distance between the null hypothesis value and the true population parameter value. For proportion, the effect is

$$|p - p_0|$$

- The larger the effect size, the larger the power of the test.

Note:

There is a relationship between confidence intervals and two-tailed hypothesis tests. We reject H_0 at the level α when the corresponding $100(1-\alpha)\%$ confidence interval does not contain the null hypothesis value p_0.