A Statistical Test of Hypothesis

- "vaccine is 50% effective" ← how to confirm, how to trust?

statistical test of hypothesis
5 parts

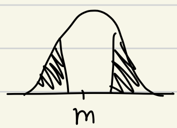
1. null hypothesis : H_0 ⟵ contradicting, mutually exclusive
2. alternative hypothesis : H_a ⟵
3. test statistic & its p-value
4. significance level and rejection region
5. conclusion

H_a is what the researcher supports
H_0 is it's contradicting hypothesis } ↗ always population parameters

- you start by assuming H_0 is true.

to prove H_a is true, you prove H_0 as false using the sample data to decide whether the evidence favours H_a rather than H_0.
Conclude by either reject H_0 or holding that it is true

$\mu \neq m$ ← value ⟶ 2 tailed test of hypothesis



$\mu < m$ ⟶ 1 tailed test of hypotheses

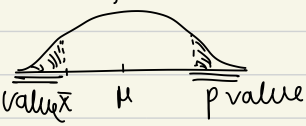
< : left tailed
> : right tailed

test statistic ($\bar{x}$) : number calced from sample data based on best estimator for the parameter to be tested.

p-value : a probability calculated using the test statistic

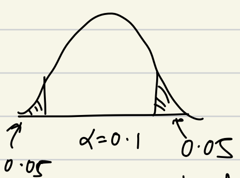
P (getting sample stat. | H_0 is true)

test statistic, not mean.



will be given (set before experiment) → α = P (falsely rejecting H_0) : level of significance

↳ represents the maximum tolerable risk of incorrectly rejecting H_0



if p-value < α ⇒ reject H_0

else ⇒ do not reject H_0

∴ shaded becomes rejection zone

Large Sample Test About a Population Mean

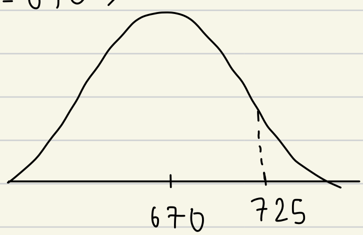
eg. question.

The average weekly earnings for female social workers is \$670. Do men in the same positions have average weekly earnings that are higher than those for women? A random sample of $n = 40$ male social workers showed $\bar{x} = \$725$ and s=\$102.

Test the appropriate hypothesis using $\alpha = 0.01$

$H_0 : \mu = 670$ ← μ = avg earnings of men
$H_a : \mu > 670$

$n = 40$, $\bar{x} = 725$, $s = 102$
p-value $= P(\bar{x} \geq 725 \mid \mu = 670)$



$\because n > 30$ can approximate :

$$z = \frac{725 - 670}{s/\sqrt{n}} = 3.41$$

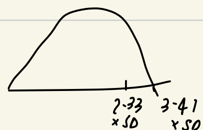
$3.41 > Z_{0.01}$ (2.33) & $p < \alpha$, $0.00032 < 0.01$

$\therefore$ reject H_0, approve H_a

from table pick $Z_{0.01}$
$\because$ one sided.
if 2 sided : $Z_{\alpha/2}$

if p value in sd format $> Z_\alpha$ → reject
if p value in z format $< \alpha$ → reject

probability that this is wrong : $\alpha = 0.01$



more about the P-value :

P-value = observed significance level of a statistical test is the smallest value of α for which H_0 can be rejected.
It is the actual risk of committing a Type 1 error, If H_0 is rejected based on the observed value of the test statistic.

P-value measures the strength of evidence against H_0

small P-value $\Rightarrow$ test stat is far from hypothesised value of μ $\longrightarrow$ strong evidence that H_0 is false and should be rejected.

so if p-value $\leq \alpha \longrightarrow$ reject H_0, results are statistically significant at level α.

The quality control manager wants to know whether the daily yield at a local chemical plant—which has averaged 880 tons for the last several years—has changed in recent months.

A random sample of 50 days gives an average yield of 871 tons with a standard deviation of 21 tons. Calculate the *p*-value for this two-tailed test of hypothesis. Use the *p*-value to draw conclusions regarding the statistical test.

$H_0 : \mu = 880$

$H_a : \mu \neq 880$

↑ 2 sided.

$n = 50$

$\bar{x} = 871$

$S = 21$

$$z = \frac{871 - 880}{21/\sqrt{50}} = -3.03$$

p value $:= P(|z| > 3.03)$

↳ 2×0.00122

$= 0.00244$

if $p \leq \alpha$, reject H_0

$p < 0.01$ → reject, highly significant

$0.01 < p < 0.05$ → reject, statistically significant

$0.05 < p < 0.1$ → not rejected, tend to statiscal significance

$p > 0.1$ → not rejected, not statistically significant

Standards set by government agencies indicate that Americans should not exceed an average daily sodium intake of 3300 milligrams (mg).

To find out whether Americans are exceeding this limit, a sample of 100 Americans is selected, and the mean and standard deviation of daily sodium intake are found to be 3400 mg and 1100 mg, respectively. Use $\alpha = .05$ to conduct a test of hypothesis.

$H_0 : \mu \leq 3300$
$H_a : \mu > 3300$

$n = 100$
$\bar{x} = 3400$
$s = 1100$
$\alpha = 0.05$
one-tailed.

$z : 0.9 < 1.645$ → dont reject

$0.18406 > 0.05$ prob of test give hypo is true is above the limit ∴ not rejected.

Types of errors

Decision	Null Hypothesis True	Null Hypothesis False
Reject H_0	Type 1 error	correct
Accept H_0	Correct	Type II error

Type I : Reject H_0 when it is true
Type II : Accept H_0 when it is false

Prob (Type I) $= \alpha$
Prob (Type II) $= \beta$

Large-Sample Test of Hypothesis for Difference Between Two Population Means

- if n is large $(\bar{x}_1 - \bar{x}_2)$ follows an approx. normal distribution w/ mean $(\mu_1 - \mu_2)$
  std error : $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
  estimated by $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$H_0 : (\mu_1 - \mu_2) = D_0$

H_a : one tailed : $(\mu_1 - \mu_2) \gtrless D_0$
two tailed : $(\mu_1 - \mu_2) \neq D_0$

Test statistic : $z \approx \dfrac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

Rejection region : Reject H_0 when (one) $z > z_\alpha$ or (two) $z > z_{\alpha/2}$ / $z < -z_{\alpha/2}$

or $p < \alpha$

assuming $n_1 \geq 30$, $n_2 \geq 30$

To determine whether car ownership affects a student's academic achievement, random samples of 100 car owners and 100 nonowners were drawn from the student body.

The grade point average for the $n_1 = 100$ nonowners had a average and variance equal to $\bar{x}_1 = 2.70$ and $s_1^2 = 0.36$ while $\bar{x}_2 = 2.54$ and $s_2^2 = 0.40$ for the $n_2 = 100$ car owner Do the data present sufficient evidence to indicate a difference in the mean achievements between car owners and nonowners? Using $\alpha = 0.05$ for the test.

$n_1 = n_2 = 100$

$\bar{x}_1 = 2.70$

$s_1^2 = 0.36$

$\bar{x}_2 = 2.54$

$s_2^2 = 0.4$

$\alpha = 0.05$

$H_0 : \mu_1 - \mu_2 = 0$

$H_a : (\mu_1 - \mu_2) \neq 0$ ← 2 tailed

$$z = \frac{(2.70 - 2.54) - 0}{\sqrt{\frac{0.76}{100}}} = 1.835$$

$z_{\alpha/2} = 1.96$ $\qquad$ $z < z_{\alpha/2} \rightarrow$ can't reject

p value $= 2 \times 0.329 = 0.658 > 0.05$

much greater exceeds significance.

Hypothesis Testing + Confidence Intervals

If hypothesized values lie outside of the confidence limits, the null hypothesis is rejected at the α level of significance

← from prev example

95% confidence interval for diff in academic achievements b/w car owners and nonowners

↳ $(\bar{x}_1 - \bar{x}_2) \pm 1.96 \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \rightarrow 0.16 \pm 0.17$
$= -0.01 < (\mu_1 - \mu_2) < 0.33$

∵ $H_0 : \mu_1 - \mu = 0$ is included in interval, you can't reject

↑

not enough evidence to indicate a difference

- gives max & min value of statistic w/ probability that it lies between those values

Statistical Significance and Practical Importance

significant ⇒ results could not have occurred by chance

Large-Sample Test of Hypothesis for a Binomial Proportion

- if n is large, $\hat{p}$ follows an approx normal dist (CLT) w/ mean p & SE

  $\hookrightarrow = \sqrt{\frac{pq}{n}}$

- $H_0 : p = p_0$

  $H_a : p \gtrless p_0$ (one tail) , $p \neq p_0$ (two tailed)

- test statistic : $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$ $\hat{p} = \frac{x}{n}$

- rejection region : Reject H_0 when

  $z > z_\alpha$ or $z < -z_\alpha$ (one)

  $z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$ (two)

  or $p < \alpha$

- satisfies assumptions of a binomial experiment & n is large enough so that CLT applies

Regardless of age, about 20% of American adults participate in fitness activities at least twice a week. Does this percentage decrease as people get older? In a local survey of n = 100 adults over 40 years old, a total of 15 people indicated that they participated in a fitness activity at least twice a week. Do these data indicate that the participation rate for adults over 40 years of age is significantly less than the 20% figure? Calculate the *p*-value and use it to draw the appropriate conclusions. Using $\alpha = 0.1$

$H_0 : p = 0.2$
$H_a : p < 0.2$

$n = 100$
$\hat{p} = 0.15$
$\alpha = 0.1$
$SE = 0.04$

$z = 1.25$
$p = 0.10565 > \alpha$ $\therefore$ can't reject

insufficient evidence

Large-Sample Test of Hypothesis for the Difference b/w 2 Binomial Proportions

- $H_0 : (p_1 - p_2) = p_0$
  $H_\alpha : (p_1 - p_2) \gtrless p_0$

- $$z = \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

  $$= \frac{(\hat{p}_1 - \hat{p}_2) - p_0}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}}$$

  $p_1 = \frac{x_1}{n_1}$ $\quad p_2 = \frac{x_2}{n_2}$

  $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$ $\quad SE = \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}$

- rejection region

  $z > z_\alpha$ (one tail)

  $z > z_{\alpha/2}$ (two tails)

- assume : samples are random + independent from 2 binomial populations
  n_1 & n_2 are large enough so the sampling distribution of $(\hat{p}_1 - \hat{p}_2)$

  $n_1\hat{p}_1 ,\ n_1\hat{q}_1 ,\ n_2\hat{p} ,\ n_2\hat{q}_2 > 5$

The records of a hospital show that 52 men in a sample of 1000 men versus 23 women in a sample of 1000 women were admitted because of heart disease.

Do these data present sufficient evidence to indicate a higher rate of heart disease among men admitted to the hospital? Use $\alpha = 0.05$.

$H_0: \quad p_1 - p_2 = 0$

$H_a: \quad p_1 - p_2 > 0$

$\hat{p}_1 = 52/1000$ $\qquad \hat{p}_1 - \hat{p}_2 = 29/1000$

$\hat{p}_2 = 23/1000$

$SE = 0.00847$

$z = 3.42$ $\qquad z > z_\alpha \rightarrow$ reject!

$z_\alpha = 1.645$

how much higher is the proportion

use confidence interval

use 95% one sided confidence interval :

$\alpha_{0.05} = 1.645$

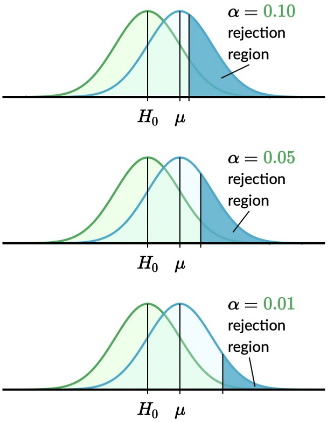
$(\hat{p}_1 - \hat{p}_2) - 1.645(SE) = 0.029 - 0.014 = 0.015$

$\therefore$ proportion of men is 1.5% higher than women.

Using a higher significance level increases the probability of a Type I error, but decreases the probability of a Type II error (which is more dangerous in this setting).

Since we are assuming that the null hypothesis is false, the correct conclusion would be to reject the null hypothesis.

Rejecting a false null hypothesis is more likely to happen with a higher significance level and less likely to happen with a lower significance level, since rejecting with lower significance requires our sample result to be farther away from the null hypothesis.



The probability of a Type II error would increase and the power of the test would decrease.

Lowering α decreases the probability of a Type I error, increases the probability of a Type II error, and decreases power.