Elemi statisztika fizikusoknak

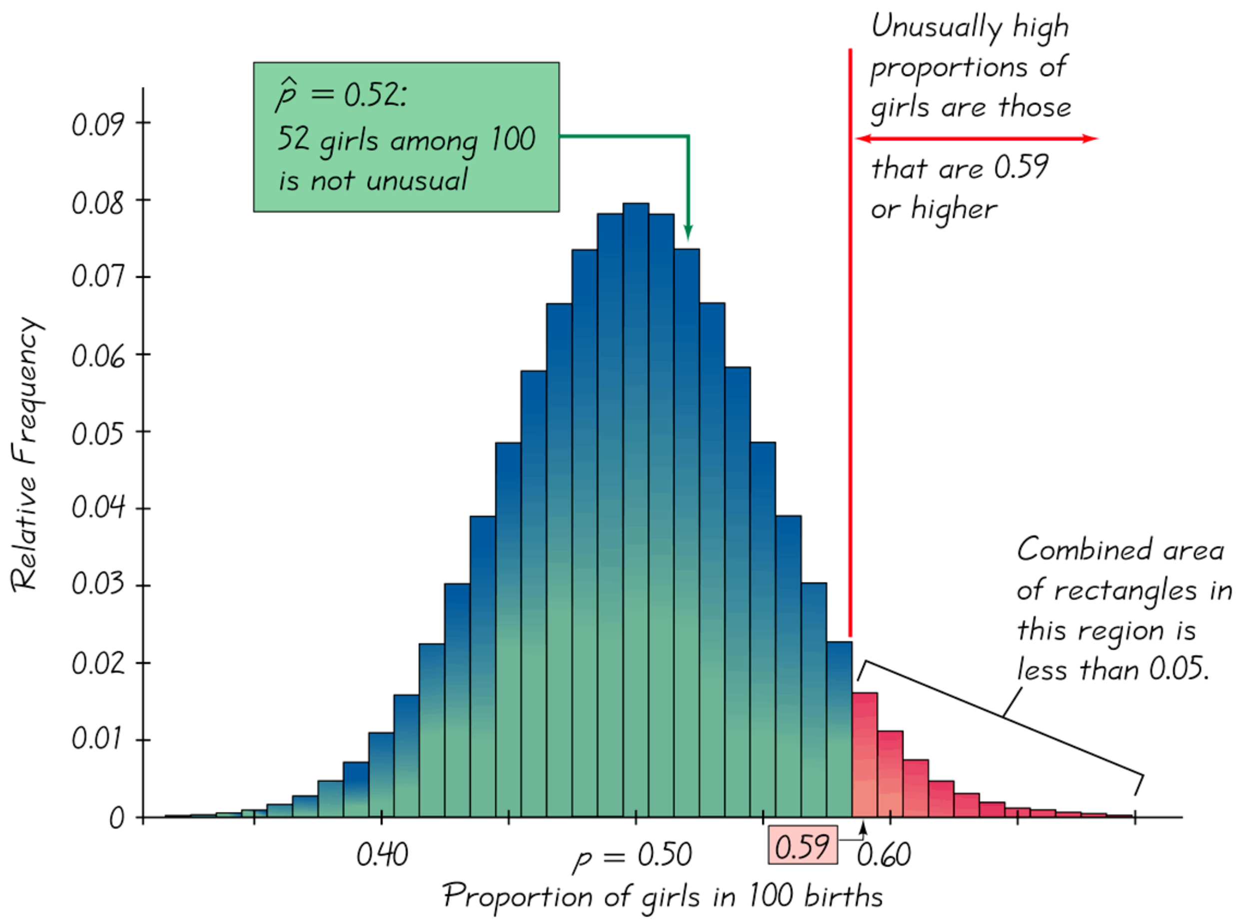
Pollner Péter

Biológiai Fizika Tanszék

pollner@elte.hu

A ritka esemény szabály a statisztikában

Ha, adott feltevések mellett egy bizonyos esemény valószínűsége kicsi, de mi mégis megfigyeljük egy ilyen esemény bekövetkezését, akkor arra a konklózióra jutunk, hogy a feltevés nem igaz.



8-1. ábra

Nem utasítjuk el a véletlent, mint elfogadható magyarázatot. Azt a konklúziót vonjuk le, hogy a „Gender Choice” által elért hatás nem szignifikánsan nagyobb, mint amit véletlenül is kaphatnánk.

A formális hipotézis teszt összetevői

Kulcsfogalmak

Ebben a fejezetben a hipotézis vizsgálat elemi összetevőit mutatjuk be, amelyeket majd a további fejezetekben használunk fel. A következő fogalmakat kell megértenünk:

- **null hipotézis**
- **alternatív hipotézis**
- **teszt statisztika**
- **kritikus tartomány**
- **szignifikancia szint**
- **krtitikus érték**
- ***P*-érték**
- **Első és másodfajú hiba (Type I and II error)**

Null hipotézis: H_0

- A null hipotézis (jelölés: H_0) egy állítás a populáció valamilyen paraméter értékéről (mint arány, átlag vagy szórás) miszerint az egyenlő valamilyen feltételezett (hipotetikus) értékkel.
- A null hipotézist közvetlenül tesztelhetjük.
- Vagy elutasítjuk a H_0 hipotézist vagy nem tudjuk elutasítani a H_0 hipotézist.

Megjegyzés a mi saját feltevésünk (hipotézisünk) kialakításával kapcsolatban

Ha valamilyen vizsgálatot végzünk és a hipotézis tesztelést akarjuk használni a saját feltevésünk alátámasztására, akkor azt úgy kell megfogalmazni, hogy a saját feltevésünk legyen az alternatív hipotézis.

Alternatív hipotézis: H_1

- Az alternatív hipotézis (jelölés H_1 vagy H_a vagy H_A) egy állítás, ami szerint a paraméter értéke valamilyen módon különbözik a nulla hipotézistől.

- Az alternatív hipotézis szimbolikus kifejezése az alábbi szimbólumokat kell, hogy tartalmazza: $\neq$, $<$, $>$.

Megjegyzés H_0 és H_1 megválasztásának lépéseiről

1. Azonosítsd a tesztelendő hipotézist és írd le szimbolikusan.
2. Add meg azt a szimbolikus alakot, aminek akkor kell igaznak lennie, ha az eredeti hipotézis hamis.
3. A fenti kettő közül az legyen a null hipotézis, amiben = szerepel. Legyen az alternatív hipotézis, amiben <, > vagy ≠ szerepel.

Példa: **Azonosítsd a null és az alternatív hipotézist az előbbiek alapján!**

a) Azon vezetők aránya, akik bevallják, hogy néha piros lámpán is áthaladnak nagyobb, mint 0.5.

b) A profi kosarasok átlag magassága legfeljebb 213cm.

c) A színészek IQ értékeinek szórása 15.

Példa: folyt.

a) **Azon vezetők aránya, akik bevallják, hogy átmennek a piroson nagyobb, mint 0.5.**

Az 1. lépésben: kifejezzük a feltevést szimbolikusan $p > 0.5$.

A 2. lépésben: látjuk, ha $p > 0.5$ hamis, akkor $p \leq 0.5$ lesz igaz.

A 3. lépésben: látjuk, hogy a $p > 0.5$ kifejezés nem tartalmaz egyenlőség jelet, ezért legyen az alternatív hipotézis (H_1) $p > 0.5$, és a null hipotézis (H_0) legyen $p = 0.5$.

Példa: folyt.

b) A profi kosarasok átlag magassága legfeljebb, 213cm.

Az 1. lépésben: szimbolikusan kifejezzük $\mu \leq 213$.

A 2. lépésben: látjuk, ha $\mu \leq 213$ hamis, akkor $\mu > 213$ igaz.

A 3. lépésben: látjuk, hogy $\mu > 213$ nem tartalmaz egyenlőség jelet, ezért ez lesz az alternatív hipotézis (H_1: $\mu > 213$), és H_0 lesz $\mu = 213$.

Példa: folyt.

c) A színészek IQ értékeinek szórása 15.
Az 1. lépésben: kifejezzük az állítást szimbolikusan $\sigma = 15$.
A 2. lépésben: látjuk, hogy ha $\sigma = 15$ hamis, akkor $\sigma \neq 15$ igaz lesz.
A 3. lépésben: az alternatív hipotézis H_1 lesz $\sigma \neq 15$, és H_0 lesz $\sigma = 15$.

Teszt statisztika

A teszt statisztika egy olyan számérték, aminek segítségével döntést tudunk hozni a null hipotézisről. A minta statisztika értékéből képezzük annak a feltevésével, hogy a null hipotézis igaz.

Teszt statisztika - képletek

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Teszt statisztika az arányra

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

Teszt statisztika az átlagra

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Teszt statisztika a varianciára

Feladat: Egy n = 880 véletlenül kiválasztott vezetőt megkérdezve 56%-uk (vagyis p = 0.56) mondta, hogy néha áthajt a piros jelzésen. Keressük meg a teszt statisztika értékét ahhoz a feltevéshez (hipotézishez), miszerint a vezetők többsége elismeri, hogy néha átmegy a piroson. (Majd a 8-3. fejezetben lesz néhány feltétel, aminek teljesülnie kell, most tegyük fel, hogy ezek rendben vannak.)

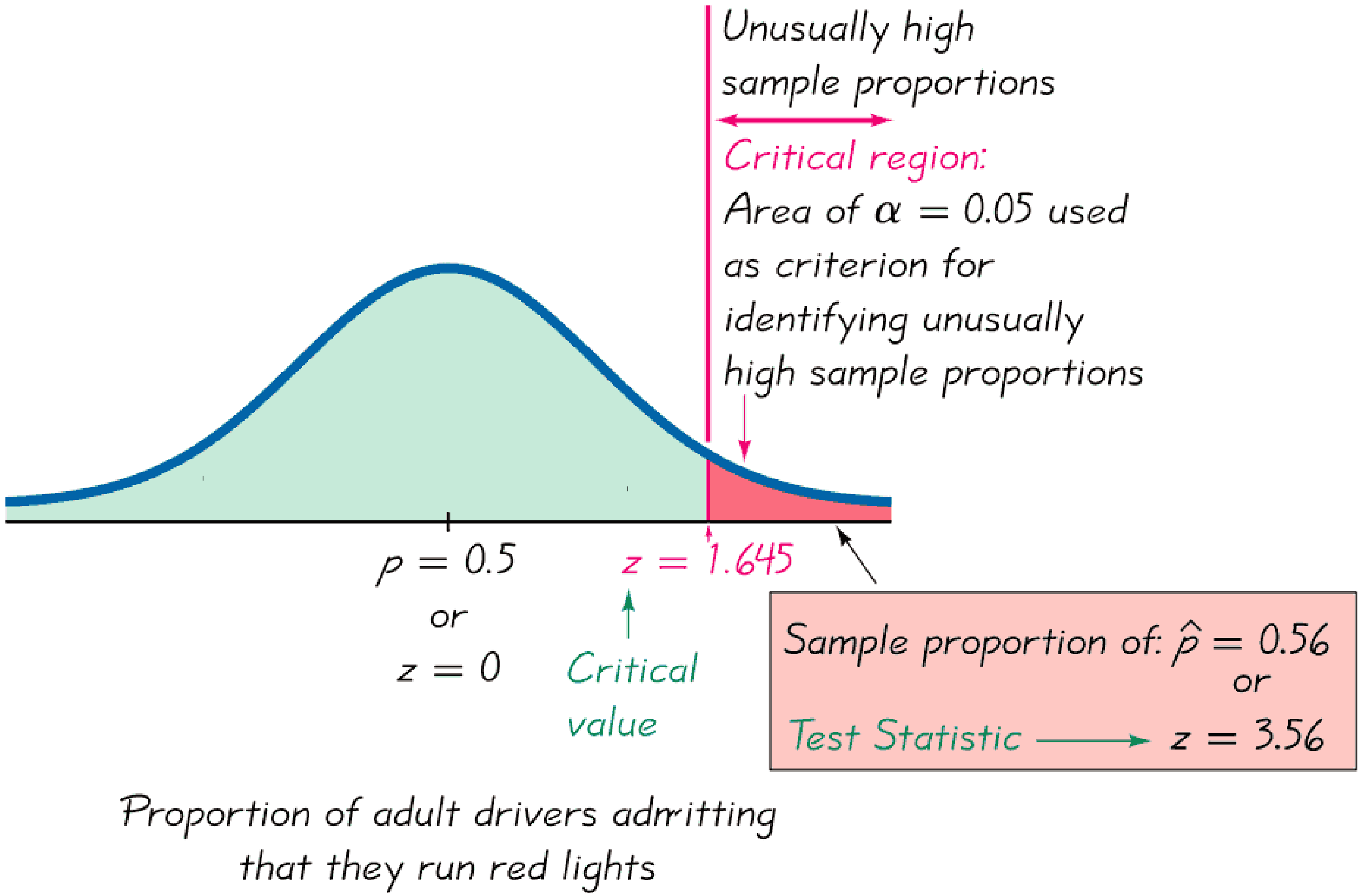
Megoldás: Az előző példában beláttuk, hogy ennek a feltevésnek az ellenőrzéséhez a H_0: $p = 0.5$ null hipotézis és a H_1: $p > 0.5$ alternatív hipotézis tartozik. Mivel azzal a feltevéssel dolgozunk, hogy a null hipotézis igaz a $p = 0.5$ értékkel, a következő teszt statisztikát kapjuk:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.56 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{880}}} = 3.56$$

Interpretáció: Tudjuk az előző fejezetekből, hogy a z=3.56 érték kivételesen nagy. Úgy tűnik, hogy azon túl, hogy az érték “több mint fél”, a minta erdmény (56%) **szignifikánsan** több mint 50%.

Ld. a következő ábrát.

Kritikus tartomány, kritikus érték, teszt statisztika



Kritikus tartomány

A kritikus tartomány (vagy elutasítási tartomány) a teszt statisztika értékeinek az a tartománya, ami arra vezet, hogy a null hipotézist elutasítsuk. Példa rá az előző ábrán a pirosra színezett rész.

Szignifikancia szint

A szignifikancia szint (jelölés: α) az a valószínűség, amivel a teszt statisztika a kritikus tartományba esik, amikor a null hipotézis valójában igaz. Ez ugyanaz az α amit a 7-2. fejezetben vezettünk be. A szokásos választások α-ra: 0.05, 0.01 és 0.10.

Kritikus értékek

A **kritikus értékek** amik elválasztják a kritikus tartományt (ahol elutasítjuk a null hipotézist) azoktól az értékektől, ahol nem utasítjuk el. A kritikus értékek függnek a null hipotézis fajtájától, a minta eloszlástól és a szignifikancia szinttől. Ld. az előző ábrát, ahol a kritikus érték $z = 1.645$ az $\alpha = 0.05$ konfidencia szinthez tartozik.

Kétoldali, jobboldali és baloldali tesztek

Az eloszlás farkai/szélső tartományai az extrém tartományok, melyeket a kritikus értékek határolnak.

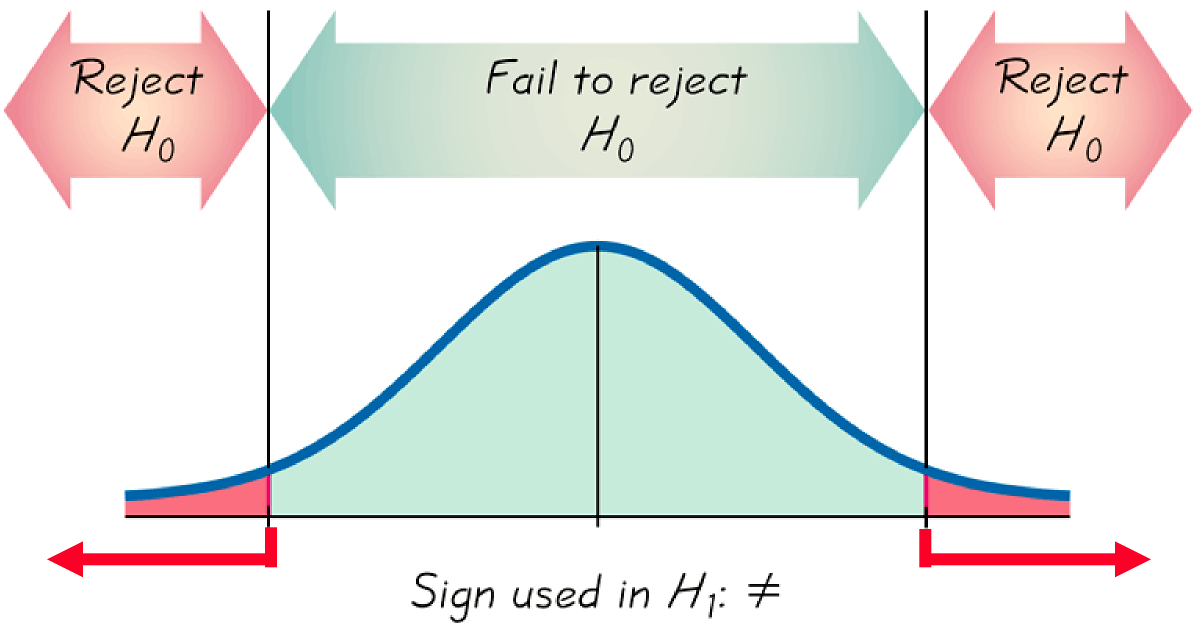
Kétoldali tesztek

H_0: $=$

H_1: $\neq$

Azt jelenti, kevesebb vagy több mint

α **egyenlően van szétosztva a két farok között**

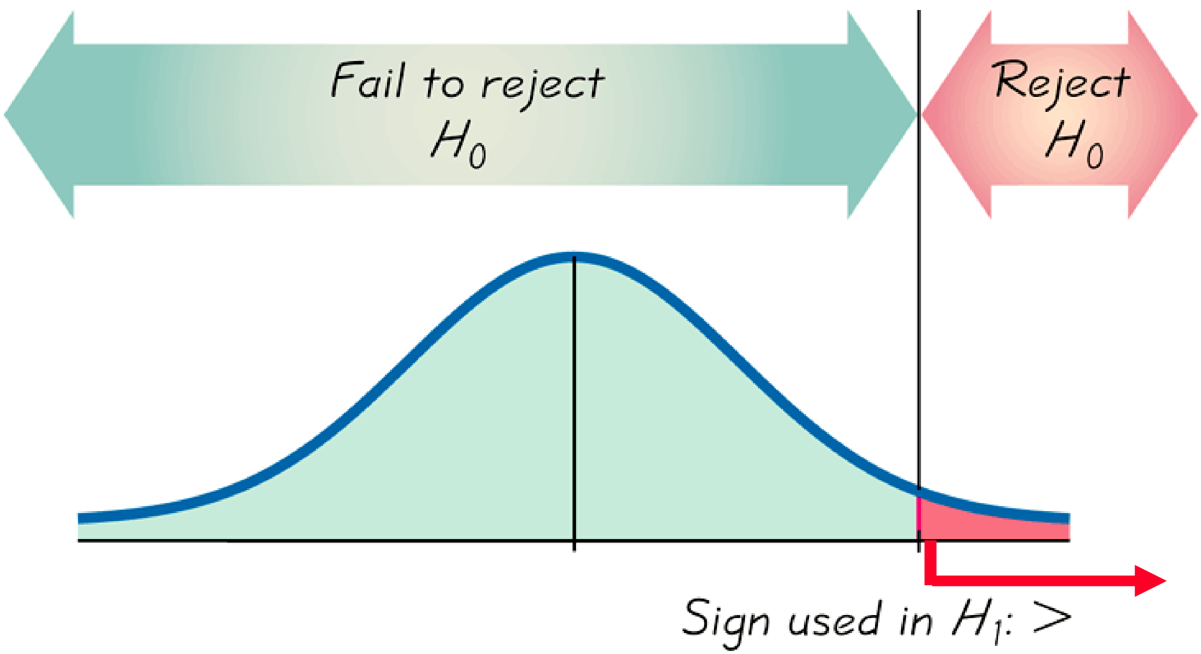


Jobboldali tesztek

H_0: =

H_1: >

Pontok jobbra vmitől

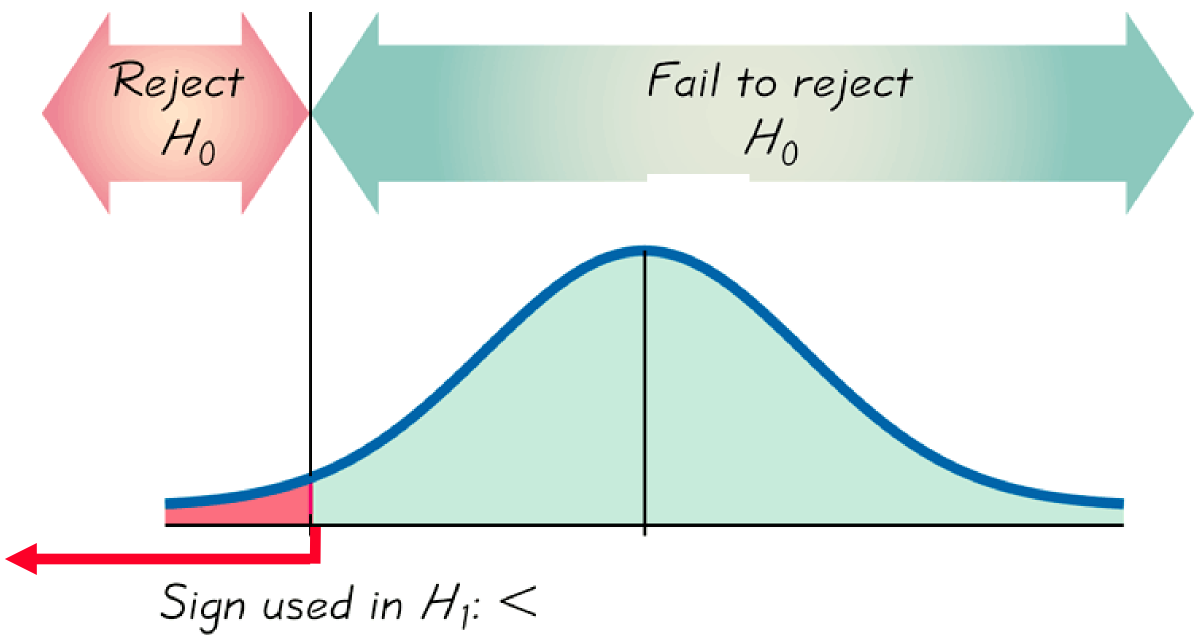


Baloldali tesztek

H_0: =

H_1: <

Pontok balra vmitől



P-érték

A *P*-érték (vagy *p*-érték vagy valószínűség érték) annak a valószínűsége, hogy a teszt statisztika olyan értéket adjon ami legalább annyira szélsőséges (extrém) mint az az érték amit a mintánkból kaptunk, azzal a feltevéssel, hogy a null hipotézis igaz. A null hipotézist elvetjük, ha a *P*-érték nagyon kicsi, mint pl. 0.05 vagy kevesebb.

A hipotézis tesztelés eredménye

Mindig a null hipotézist teszteljük. A kezdeti konklúzió mindig az alábbiak valamelyike:

1. Elvetjük a null hipotézist.

2. Nem tudjuk elvetni a null hipotézist.

Döntési kritériumok

Tradicionális módszer:

Elvetjük H_0-t, ha a teszt statisztika a kritikus tartományba esik.

Nem tudjuk elvetni H_0-t, ha a teszt statisztika nem esik a kritikus tartományba.

Döntési kritériumok - folyt

***P*-érték módszer:**

Elvetjük H_0-t ha a *P*-érték $\leq \alpha$ (ahol α a szignifikancia szint, mint pl. 0.05).

Nem tudjuk elvetni H_0-t, ha a *P*-érték $> \alpha$.

Döntési kritériumok - folyt

Egy másik lehetőség:

A szignifikancia szint megadása helyett, egyszerűen megkeressük a *P*-értéket, és a döntést az olvasóra hagyjuk.

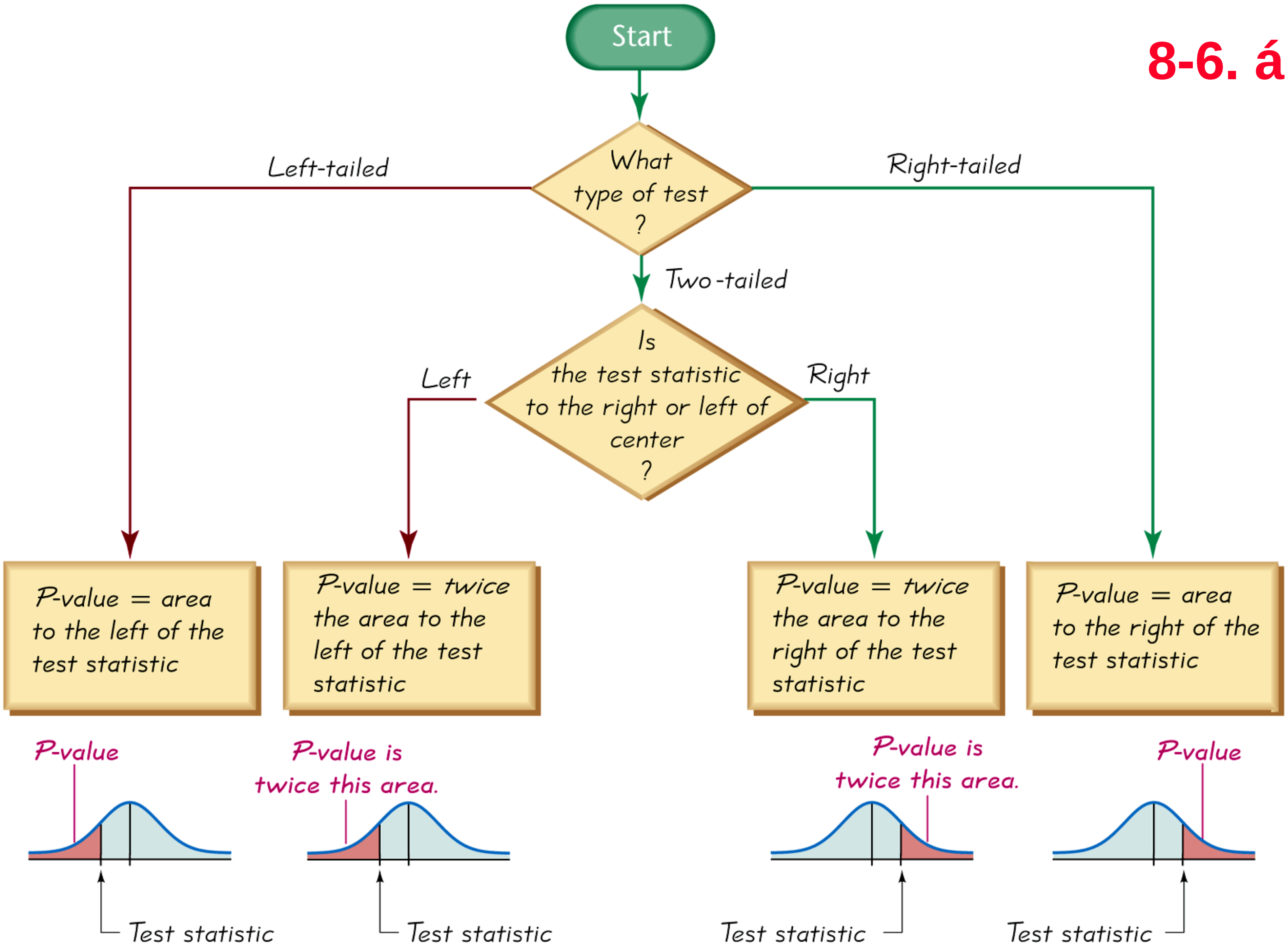
Döntési kritérium – folyt.

Konfidencia intervallum:

Mivel a konfidencia intervallum becslés tartalmazza a paraméter populációbeli értékét, utasítsuk el azokat a feltevéseket, melyek szerint a populáció paramétere a konfidencia intervallumon kívül esik.

A *P*-értékek megtalálásának ábrája

8-6. ábra



Példa: ***P*-érték kiszámítása. Először határozzuk meg, hogy az adott esetben jobboldali, baloldali vagy kétoldali tesztet végzünk-e, azután keressük meg a *P*-értéket és adjuk meg a null hipotézissel kapcsolatos konklúziót.**

a) Az $\alpha = 0.05$ szignifikancia szintet használjuk annak a feltételezésnek a tesztelésére, hogy $p > 0.25$, és a minta adatok egy $z = 1.18$ értékű teszt statisztikát adnak.

b) Az $\alpha = 0.05$ szignifikancia szintet használjuk annak a feltételezésnek a tesztelésére, hogy $p \neq 0.25$, és a minta adatok egy $z = 2.34$ értékű teszt statisztikát adnak.

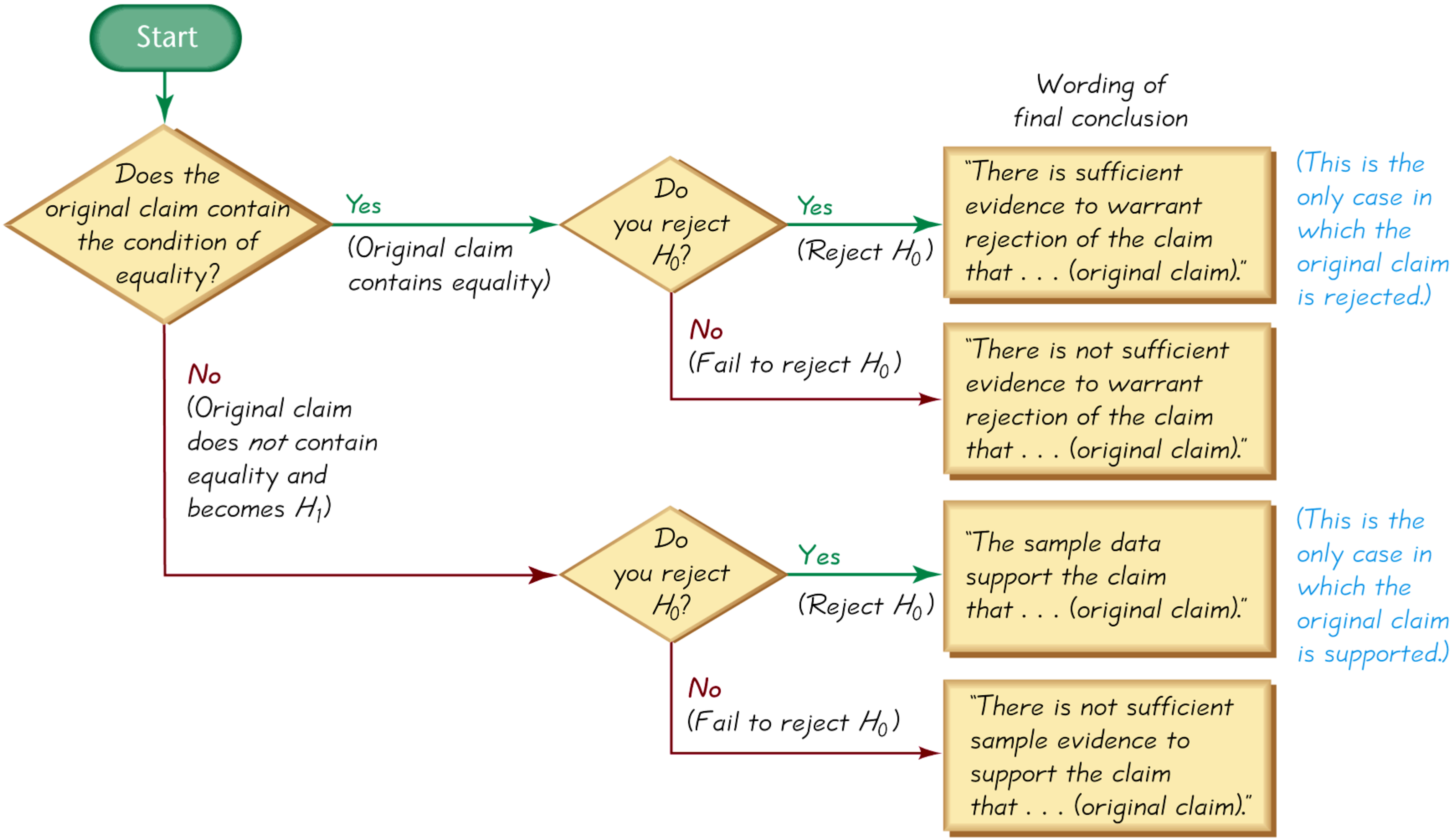
Példa: folyt.

a) A $p > 0.25$ feltevés esetén a teszt jobboldali. Mivel a teszt jobboldali, a P-érték a $z = 1.18$-től jobbra eső görbe alatti terület. Kikeresve a táblázatból ez 0.1190. A P-érték (0.1190) nagyobb mint a szignifikancia szint $\alpha = 0.05$, így nem tudjuk elvetni a null hipotézist. A P=0.1190 elég nagy, ami azt jelenti, hogy a minta érték könnyen megtörténhet véletlenül.

Példa: folyt.

b) A $p \neq$ 0.25 feltevés esetén a teszt kétoldali. Mivel a teszt kétoldali és mivel a teszt statisztika z = 2.34 a középtől jobbra esik, a P-érték kétszerese a z = 2.34-től jobbra eső területnek. A táblázatból a z = 2.34-től jobbra eső terület 0.0096, így a P-érték = 2 x 0.0096 = 0.0192. Mivel a P=0.0192 kisebb vagy egyenlő mint a szignifikancia szintünk, el kell vetnünk a null hipotézist. A kicsiny P-érték (0.0192) azt mutatja, hogy a minta eredmény valószínűleg nem a véletlen eredménye.

A végső konklúziók megfogalmazása



8-7. ábra

Elfogadni vagy nem tudni elutasítani?

- Bizonyos könyvekben azt mondják “elfogadjuk a null hipotézist.”
- Nem tudjuk bizonyítani a null hipotézist.
- A minta bizonyítékok nem elég erősek ahhoz, hogy elutasítsuk (olyan mint amikor nincs elég bizonyíték, hogy elítéljék a gyanúsítottat).

I. fajú hiba

- Egy I. fajú hiba az, amikor hibás módon elutasítjuk a null hipotézist, amikor az igaz.

- Az α (alfa) szimbólummal jelöljük az I. fajú hiba valószínűségét.

II. fajú hiba

- Egy **II. fajú hiba** az, amikor nem utasítjuk el a null hipotézist akkor, amikor az nem igaz.

- A β (béta) szimbólummal jelöljük a II. fajú hiba valószínűségét.

Példa: **Tegyük fel, hogy hipotézis tesztelést végzünk a $p > 0.5$ feltevéssel kapcsolatban. A null és az alternatív hipotézis a következő: H_0: $p = 0.5$, és H_1: $p > 0.5$.**

a) **Azonosítsuk az I. fajú hibát.**
b) **Azonosítsuk a II. fajú hibát.**

Példa: folyt.

a) Az I. fajú hiba az, amikor elvetjük az igaz null hipotézist: Ha úgy látjuk, hogy elég kicsi a valószínűsége p = 0.5-nak, miközben a valóságban p = 0.5.

Példa: folyt.

b) A II. fajú hiba az, amikor nem vetjük el a null hipotézist, miközben az nem igaz: Nem utasítjuk el p = 0.5-öt (és ezért nem támogatjuk a p > 0.5-öt), miközben a valóságban p > 0.5.

Első és másodfajú hibák

Table 8-1 Type I and Type II Errors			
		True State of Nature	
		The null hypothesis is true	The null hypothesis is false
Decision	We decide to reject the null hypothesis	**Type I error** (rejecting a true null hypothesis) α	Correct decision
	We fall to reject the null hypothesis	Correct decision	**Type II error** (failing to reject a false null hypothesis) β

Az I. és II. fajú hibák kontrollálása

- **Minden adott α esetén, a minta elemszám *n* növelése a β csökkenését okozza.**

- **Minden fix minta elemszám *n* esetén α csökkenése β növekedését okozza. Fordítva, α növelése β csökkenésére vezet.**

- **Ha α és β együttes csökkenését akarjuk elérni, akkor a minta elemszámot kell növelnünk.**

Definíció

A **hipotézis teszt erőssége** az $(1 - \beta)$ valószínűség érték, ami a helytelen null hipotézis elutasításának valószínűsége. Egy adott α szignifikancia szint és adott olyan másik populáció paraméter esetén számíthatjuk ki, ami a null hipotézisbeli érték alternatívája. Azaz a hipotézis teszt erőssége egy igaz alternatív hipotézis támogatásának valószínűsége.

Összefoglalás

Ebben a fejezetben megvitattuk:

- **Null és alternatív hipotézis.**
- **Teszt statisztika.**
- **Szignifikancia szintek.**
- ***P*-értékek.**
- **Döntési kritériumok.**
- **Első és másodfajú hibák.**
- **A hipotézis teszt ereje.**

8-3. fejezet
Az arányra vonatkozó feltevés tesztelése

Kulcsfogalmak

Ebben a fejezetben a populáció arány tesztelésének teljes folyamatát ismertetjük. Felhasználjuk az előző fejezetben bevezetett fogalmakat.

Feltevések

1) Véletlen egyszerű mintavétel.

2) A binomiális elsoszlás feltételei fennállnak (5-3 fejezet).

3) Az $np \geq 5$ és $nq \geq 5$ feltételek fennállnak, így a binomiális eloszlást egy olyan normálissal közelíthetjük, aminek a paraméterei $\mu = np$ $\sigma = \sqrt{npq}$.

Jelölések

n = a kísérletek száma

$\hat{p} = \frac{x}{n}$ (minta arány)

p = populáció arány (amit a null hipotézisben használunk)

$q = 1 - p$

Az arányra vonatkozó teszt statisztika

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

P-érték módszer

Ugyanúgy, mint a 8-2 fejezetben …

Tradicionális módszer

Ugyanaz, mint a 8-2-ben …

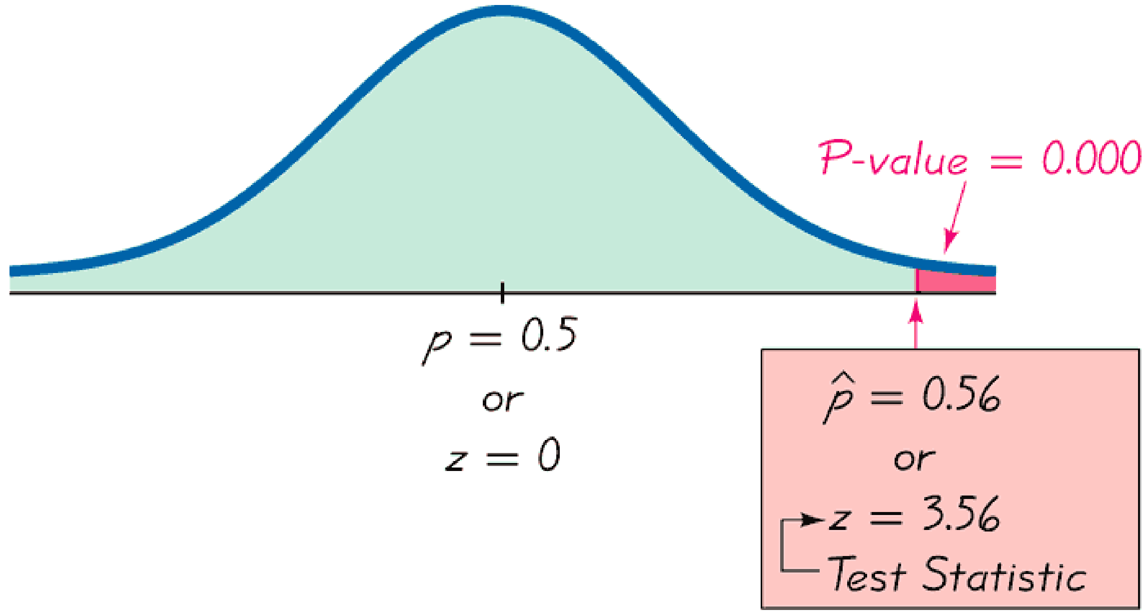
Konfidencia intervallum módszer

Ugyanaz, mint a 8-2-ben …

Példa: **880 véletlenül választott autóvezető 56%-a elismeri, hogy néha átmegy a piroson. Az a feltételezésünk, hogy a vezetők többsége néha átmegy a piroson, azaz $p > 0.5$. A minta adatok $n = 880$, $\hat{p} = 0.56$.**

$$np = (880)(0.5) = 440 \geq 5$$
$$nq = (880)(0.5) = 440 \geq 5$$

Példa: folyt.



H_0: $p = 0.5$
H_1: $p > 0.5$
$\alpha = 0.05$

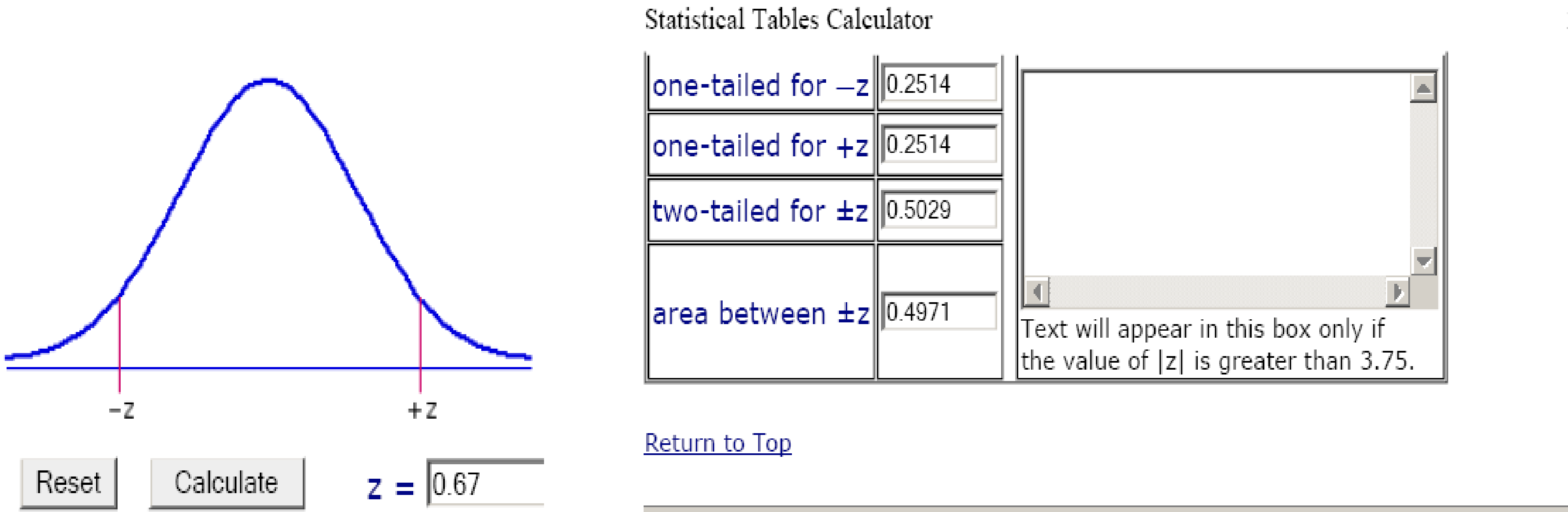
$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.56 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{880}}} = 3.56$$

P=0.0001
Mivel P<0.05=α, ezért elutasítjuk a null hipotézist.
Mivel $z>z_\alpha=1.645$, ezért elutasítjuk a null hipotézist.
Elegendő bizonyítékunk van a feltételezésünk elfogadására.

Példa: **Amikor Gregor Mendel borsó hibridizációs kísérletét végezte, az egyik kísérletben 428 zöld borsószem és 152 sárga borsószem termett. Mendel elmélete szerint a borsók ¼-e volt sárgának várható. Használjunk 0.05 szignifikancia szintű tesztet és a *P*-érték módszert, hogy teszteljük, vajon a sárga szemek aránya ¼-e vagy sem.**

Észrevétel: $n = 428 + 152 = 580$,
így $\hat{p} = 0.262$, **és** $p = 0.25$.

Segítség: http://faculty.vassar.edu/lowry/tabs.html



H_0: $p = 0.25$

H_1: $p \neq 0.25$

$n = 580$

$\alpha = 0.05$

$\hat{p} = 0.262$

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}} = \frac{0.262 - 0.25}{\sqrt{\dfrac{(0.25)(0.75)}{580}}} = 0.67$$

Mivel ez egy kétoldali teszt, a *P*-érték a kétszerese a statisztika értékétől jobbra eső területnek. P=0.502. Nincs elég bizonyítékunk, hogy elutasítsuk a null hipotézist, azaz azt, hogy a borsók ¼-e sárga.

Összefoglalás

Ebben a fejezetben megvitattuk:

- **Az arányokra vonatkozó teszt statisztikát.**
- ***P*-érték módszer.**

8-4. fejezet
Az átlagra vonatkozó feltételezés tesztelése: σ ismert

Kulcsfogalmak

Ilyet is lehet csinálni pedagógiai okokból, de gyakorlati jelentősége nincs. A következő fejezet eredményei igazak, csak s=σ-t kell feltételeznünk.

8-5. fejezet
A populáció átlagra vonatkozó feltételezés tesztelése: σ nem ismert

Kulcsfogalmak

Ebben a fejezetben a populáció átlagára vonatkozó hipotézisek vizsgálatáról lesz szó, abban az esetben, amikor σ nem adott. Ebben a fejezetben a Student *t* eloszlást használjuk.

Feltételek

1) A minta véletlen egyszerű.

2) Valamelyik, vagy mindkét feltétel igaz: A populáció normális eloszlású, vagy $n > 30$.

Teszt statisztika

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$$

P-értékek és kritikus értékek

- Táblázat, online kalkulátor, program stb.
- Szabadsági fokok száma n – 1

Példa: 13 piros M&M csoki golyót véletlenül választuk egy zacskóból, amiben 465 M&M csoki golyó van. A tömegük (grammokban) átlagosan $\bar{x}$ = 0.8635 és a szórás s = 0.0576 g.

A zacskó szerint a nettó tömeg 396.9 g, azaz az M&M csoki golyók tömege elvben 396.9/465 = 0.8535 g. Használjuk a minta adatokat és a 0.05-ös szignifikancia szintet, hogy teszteljük a gyártósor vezetőjének azon kijelentését, mi szerint a csoki golyók tömege valójában nagyobb mint 0.8535 g.

A minta n = 13 elemű és a normál q-q plot szerint normálissal közelíthető.

H_0: $\mu = 0.8535$
H_1: $\mu > 0.8535$
$\alpha = 0.05$
$\bar{x} = 0.8635$
$s = 0.0576$
$n = 13$

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{0.8635 - 0.8535}{\frac{0.0576}{\sqrt{13}}} = 0.626$$

A kritikus érték $t_\alpha = 1.782$

Mivel a teszt statisztika értéke $t = 0.626$ nem esik a kritikus tartományba, nem tudjuk elvetni a null hipotézist H_0. Nincs elegendő bizonyíték annak a feltételezésnek a támogatására, hogy az M&M csoki golyók tömege több mint 0.8535 g.

Összefoglalás

Ebben a fejezetben megvitattuk:

- **Feltételek**
- **Student *t* eloszlás.**
- ***P*-érték módszer.**

8-6. fejezet
A szórásra és a varianciára vonatkozó feltevések becslése

Kulcsfogalmak

Ebben a fejezetben a populáció szórására σ vagy varianciájára σ^2 vonatkozó feltételezés tesztelésével foglalkozunk. A módszerek támaszkodni fognak a 7-5. fejezetben bevezetett khí-négyzet eloszlásra.

Feltételek

1. Véletlen egyszerű minta.
2. A populáció normális eloszlású. (Ez egy sokkal erősebb feltétel, mint amit az átlag tesztelésekor használunk!)

Khí-négyzet eloszlás

Teszt statisztika

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma^2}$$

n = minta elemszám

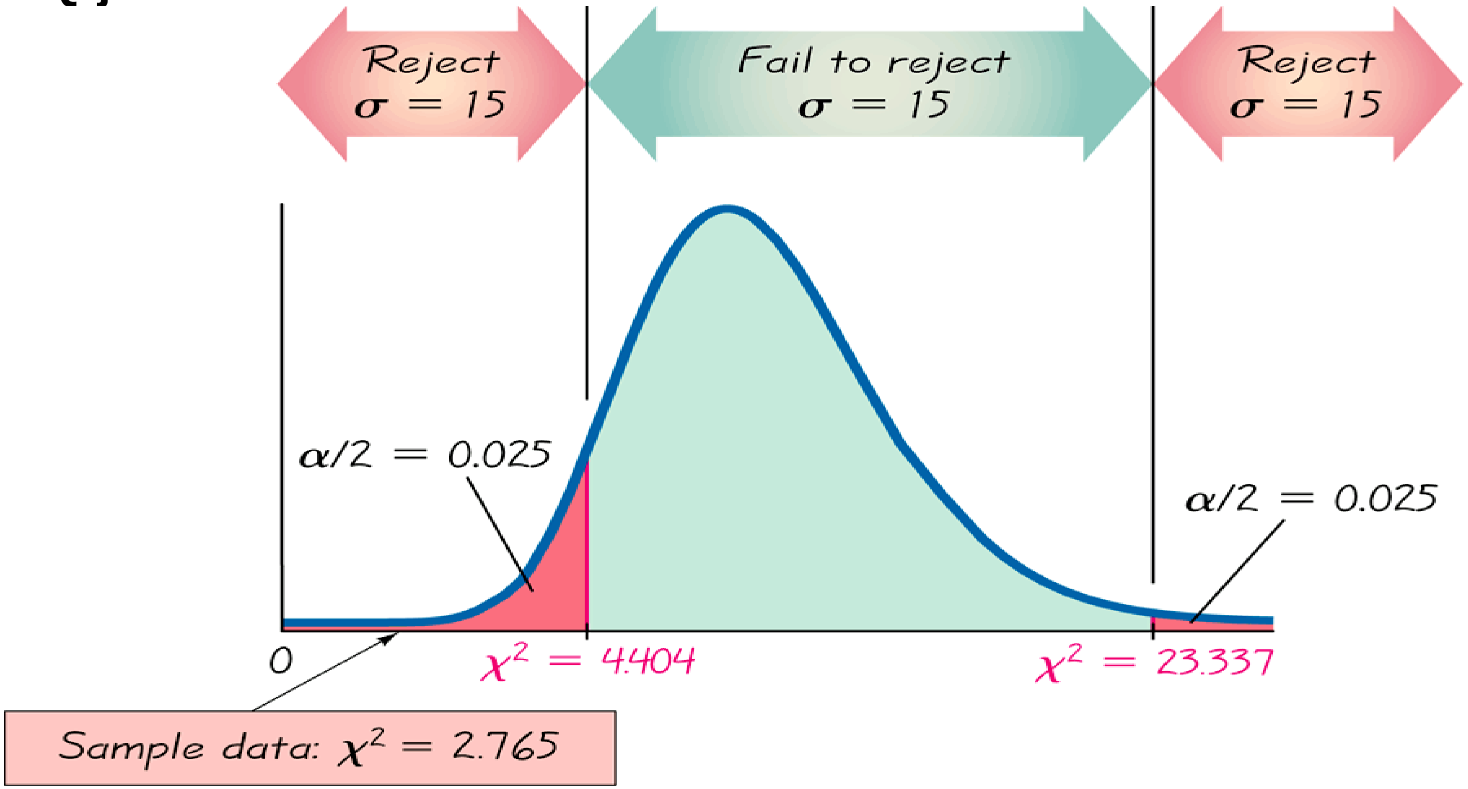
s^2 = minta variancia

σ^2 = populáció variancia

(a null hipotézisben van megadva!)

Példa: **A felnőttek egy egyszerű véletlen mintájában az IQ értékek normálisan oszlanak el 100 átlaggal és 15 szórással. Egy egyszerű véletlen 13 fizika professzorból álló mintának a szórása s = 7.2. Tegyük fel, hogy a fizika professzorok IQ-ja is normális eloszlású. Teszteljük 0.05 szignifikancia szinten azt a feltételezést, hogy a fizika professzorok IQ-jának is 15 a szórásasa $\sigma = 15$.**

H_0: $\sigma = 15$
H_1: $\sigma \neq 15$
$\alpha = 0.05$
$n = 13$
$s = 7.2$



$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(13-1)(7.2)^2}{15^2} = 2.765$$

Példa: folyt.

H_0: $\sigma = 15$
H_1: $\sigma \neq 15$
$\alpha = 0.05$
$n = 13$
$s = 7.2$

$\chi^2 = 2.765$

A kritikus értékek 4.404 és 23.337 (szabadsági fokok száma (df) = *n* – 1=12) és a táblázatban a 0.025 és 0.975 értékekhez tartoznak. Mivel a statisztika értéke a kritikus tartományba esik, el kell utasítanunk a null hipotézist. Elegendő bizonyítékunk van arra, hogy elutasítsuk azt a feltételezést, miszerint a fizika professzorok IQ-jának szórása éppen 15.

Összefoglalás

Ebben a fejezetben megvitattuk:

- **A szórásra és a varianciára vonatkozó teszteket.**
- **A teszt statisztikát.**
- **A kritikus értékeket.**