u	0	1	2	3	4	5	6	7	8	9
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	39795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	4997674									
4,0	4999683									
4,5	4999966									
5,0	499999713									

STATISTIČKE HIPOTEZE I TESTOVI

STATISTIČKA HIPOTEZA

• često se pri obradi eksperimentalnih podataka na osnovi ograničenog broja poznatih vrijednosti neke slučajne varijable treba zaključiti **proistječe li ta slučajna varijabla iz nekog teorijskog zakona raspodjele**

• zbog **ograničenog broja podataka** odgovor na to pitanje sadrži element slučajnosti i treba ga uzimati s određenom vjerojatnošću

• **statistička hipoteza** u tom slučaju predstavlja pretpostavku da ta slučajna varijabla jest raspodijeljena po određenom zakonu raspodjele

• tu hipotezu treba testirati, odnosno ustvrditi s kolikom vjerojatnošću je možemo smatrati točnom ili s kolikom vjerojatnošću ju možemo odbaciti – **statistički test** – ocjena ispravnosti postavljene hipoteze

STATISTIČKA HIPOTEZA

• neka je *x* slučajna varijabla

• postavlja se hipoteza da ta slučajna varijabla *x* **pripada** raspodjeli s poznatom funkcijom vjerojatnosti $f_0(x)$ – **nulta hipoteza (nul-hipoteza) –** $\mathbf{H_0}$

• uvodi se i alternativna hipoteza $\mathbf{H_1}$ prema kojoj slučajna varijabla pripada raspodjeli s funkcijom vjerojatnosti $f_1(x)$ i pretpostavlja se da je hipoteza $\mathbf{H_1}$ istinita ako je nulta hipoteza $\mathbf{H_0}$ neistinita

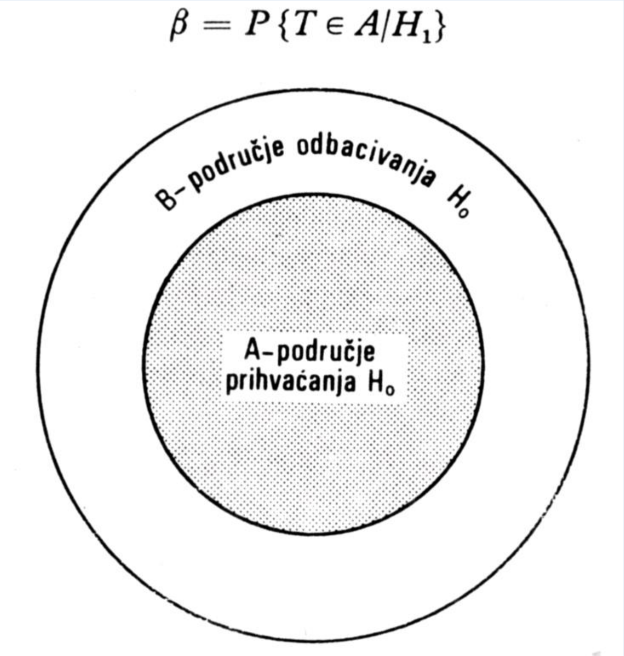
• odluka o tome koju ćemo hipotezu prihvatiti kao istinitu temelji se na osnovi određenog broja poznatih vrijednosti varijable *x*, odnosno na temelju **uzorka** koji nam je na raspolaganju

• s obzirom da našu odluku temeljimo na ograničenom uzorku, a ne na temelju populacije, postoji određena **vjerojatnost prihvaćanja pogrešne hipoteze**

STATISTIČKA HIPOTEZA

TESTIRANJE

- područje definicije x varijable podijelimo u dva disjunktna skupa



- ukoliko je točka uzorka **T** u iz dijela prostora **A** (T є A), prihvatimo $\mathbf{H_0}$ kao istinitu hipotezu
- ukoliko je točka uzorka **T** u iz dijela prostora **B** (T є B), odbacujemo $\mathbf{H_0}$ te prihvaćamo $\mathbf{H_1}$ kao istinitu hipotezu

- **A** – područje prihvaćanja hipoteze; **B** – kritična domena ili područje signifikantnosti

STATISTIČKA HIPOTEZA

POGREŠKE

- **odbacivanje** nulte hipoteze $\mathbf{H_0}$ kada je ona **istinita**, tj. prihvaćanje alternativne hipoteze umjesto nulte hipoteze naziva se ***pogreška prve vrste :*** $\alpha = P\ \{T \in B|H_0\}$

- **prihvaćanje** nulte hipoteze $\mathbf{H_0}$ kada je ona **neistinita** naziva se ***pogreška druge vrste :*** $\beta = P\ \{T \in A|H_1\}$

- **jakost testa** *(power of test)* određena je vjerojatnošću pogreške druge vrste $p = 1 - \beta$

- vjerojatnost pogreške prve vrste obično se unaprijed određuje (0,05 ili 0,01) i često se naziva i “razinom značajnosti ili signifikantnosti” *(significance level)*

  - PRIMJER BACANJA SIMETRIČNOG I ASIMETRIČNOG NOVČIĆA

STATISTIČKA HIPOTEZA

- statistički testovi koju uključuju jednu populaciju najčešće pokušavaju dati odgovore na sljedeća pitanja:
- Da li je očekivana vrijednost populacije jednaka nekoj određenoj vrijednosti?
- Da li je varijanca populacije jednaka nekoj određenoj vrijednosti?
- Da li raspodjela vjerojatnosti za populaciju ima neki određeni oblik?

STATISTIČKA HIPOTEZA

- **Da li je očekivana vrijednost populacije jednaka nekoj određenoj vrijednosti?**

Primjer:

Proizvodnja herbicida **A** standardnom metodom ima takvo iskorištenje reakcije da se u jednom ciklusu prosječno proizvede 218,0 g hrebicida uz varijanciju od 44,52 g^2. Računalnim metodama je otkriveno da bi dodatak supstance **B** trebao povećati iskorištenje reakcije. Provedeno je 10 reakcija uz prisustvo **B** te su izmjerena slijedeća iskorištenja: 235,4, 231,3, 222,7, 217,9, 214,3, 225,4, 234,2, 230,1, 221,3, 217,4. S kolikom vjerojatnošću možemo tvrditi da je dodatak tvari **B** povećao iskorištenje reakcije?

RJEŠENJE:

H_0 – srednja vrijednost uzorka (10 eksperimenata) jednaka je srednjoj vrijednosti populacije – ako je nul hipoteza točna, B ne utječe na iskorištenje reakcije

- iz priloženih podataka možemo izračunati da je srednja vrijednost uzoraka 225,0 g

- prema središnjem graničnom teoremu srednje vrijednosti raspodijeljene su po normalnoj raspodjeli

STATISTIČKA HIPOTEZA

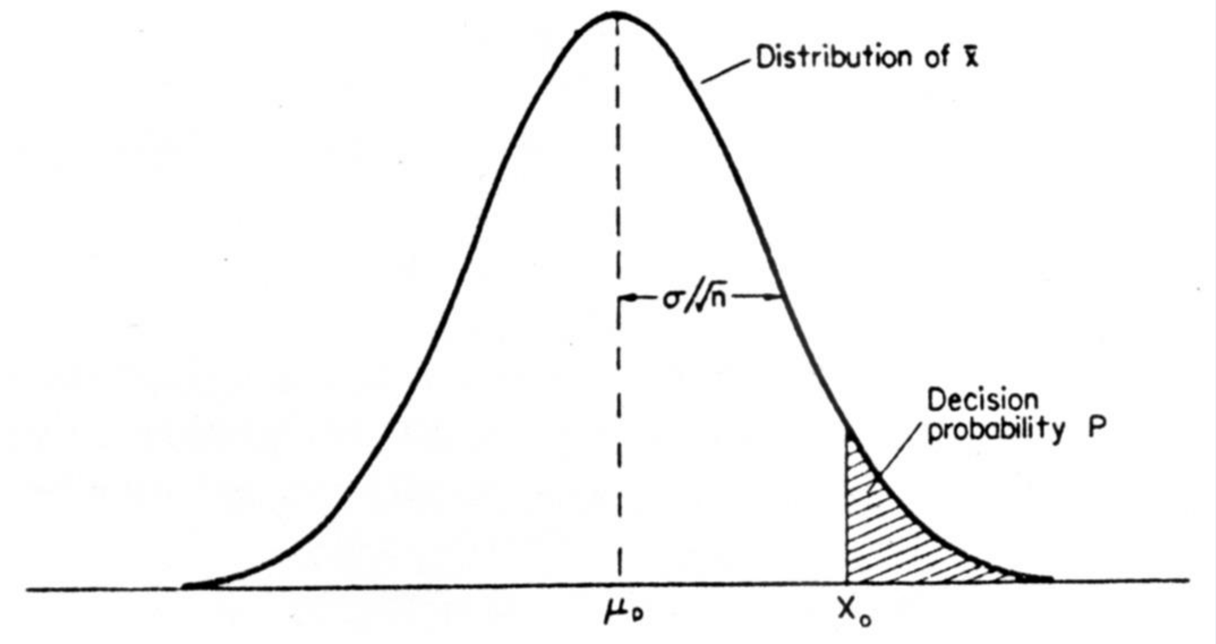
- korištenjem normalne raspodjele možemo izračunati vjerojatnost da nasumično odabrana varijabla *x* iz populacije bude veća ili jednaka srednjoj vrijednosti uzoraka

$$u = \frac{\bar{x} - \mu_p}{\sqrt{\frac{\sigma^2}{n}}} = \frac{225{,}0 - 218{,}0}{\sqrt{\frac{44{,}52}{10}}} = 3{,}32$$

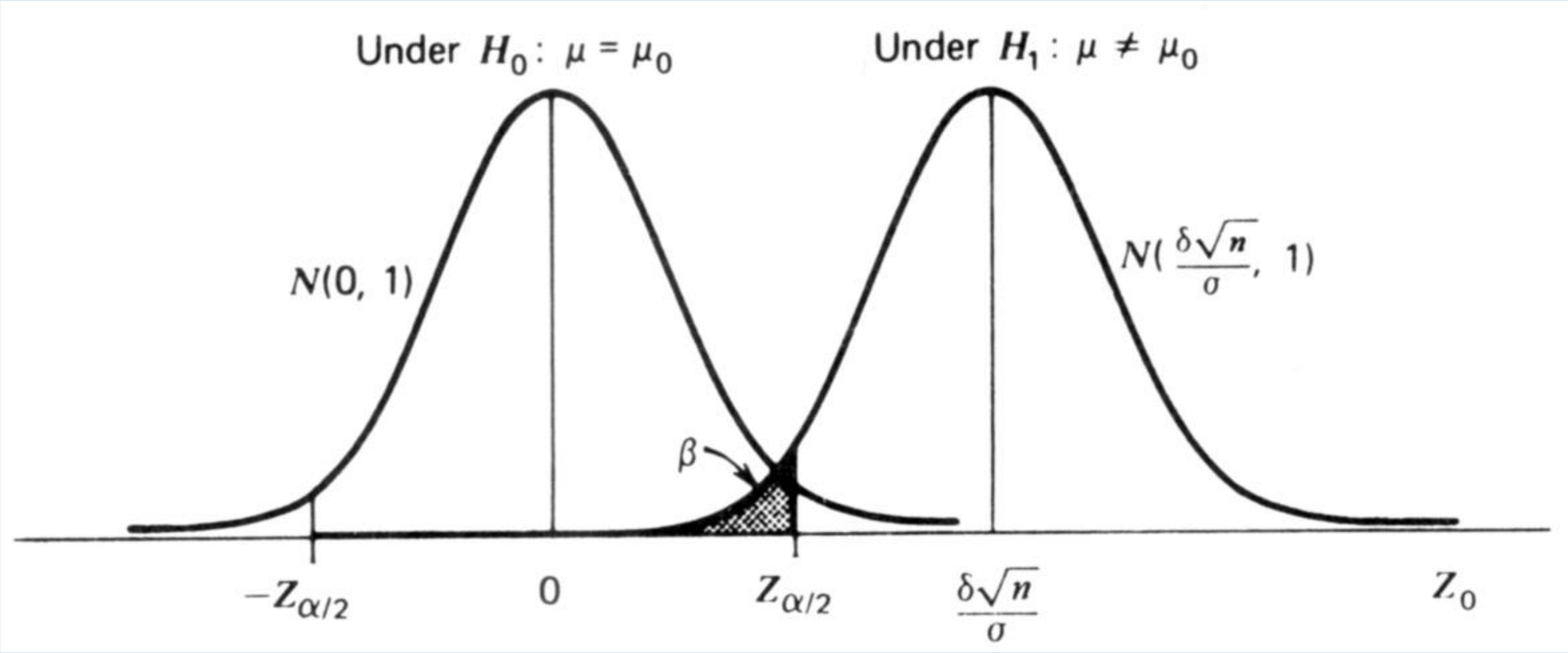
•iz tablice očitamo da je P(x ≥ 3,32 | x~ N (218,0, 44,52)) = 0,49953, što implicira da je svega 0,00047 (0,5-0,49953), odnosno 0,05% vjerojatnosti da slučajno izabrana varijabla bude veća ili jednaka srednjoj vrijednosti uzorka

• **možemo odbaciti nul hipotezu i pogreška prvog reda pri tome nam iznosi svega 0,05%**

• u navedenom primjeru koristili smo statistički test za usporedbu očekivane vrijednosti populacije u slučaju kada je varijancija bila poznata



Distribution of $\bar{X}$
$\sigma/\sqrt{n}$
Decision probability P
μ_0
X_0



Under $H_0: \mu = \mu_0$
Under $H_1: \mu \neq \mu_0$
$N(0, 1)$
$N(\frac{\delta\sqrt{n}}{\sigma}, 1)$
β
$-Z_{\alpha/2}$
0
$Z_{\alpha/2}$
$\frac{\delta\sqrt{n}}{\sigma}$
Z_0

STATISTIČKA HIPOTEZA

- **Da li je očekivana vrijednost populacije jednaka nekoj određenoj vrijednosti (varijancija je nepoznata!)?**

- u ovom slučaju koristi se sličan test u kojem se poznata vrijednost varijancije zamijeni procjenom varijancije:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

- navedeni test zove se **Studentov *t*-test** (W.S. Gosset):

$$t = \frac{\overline{x} - \mu_p}{\sqrt{\sigma_\nu^2 / n}}$$

- raspodjela funkcije *p(t)* ovisi o parametru *v*, tj. broju stupnjeva slobode pridruženom procjeni varijance

STATISTIČKA HIPOTEZA

• raspodjela funkcije *p (t)* ima oblik

$$p(t) = K\left(1 + \frac{t^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)} \qquad K = \frac{\Gamma\left[\frac{1}{2}(\nu+1)\right]}{\sqrt{\nu\pi}\,\Gamma\left(\frac{1}{2}\nu\right)}$$

• funkcija je simetrična s obzirom na *t* = 0, nešto je spljoštenija od normalne raspodjele *N (0,1)*