

Inferență statistică în ML

Cap 4. T confidence intervals. Hypotesis testing. P-values.

April 8, 2019

1 T confidence intervals

2 Hypothesis testing

3 P-values

4 Bibliografie

Intervale de confidență cu distribuția T (Student)

- am vorbit de crearea de confidență folosind CLT
- forma generală:

$$Est \pm ZQ * SE_{Est}$$

- ne vom referi la distribuții aplicabile pentru sample-uri de dimensiuni reduse
- distribuția T Student (Gosset)

$$Est \pm TQ * SE_{Est}$$

- distribuția T are cozi mai 'grele' decât distribuția normală
- intervalele de confidență asociate vor fi mai largi
- pe măsură ce sample size devine mai mare, intervalul Z tinde la intervalul T, deci folosim T interval implicit

Distribuția T a lui Gosset

- inventată de William Gosset în 1908
- “William Sealy Gosset (13 June 1876 – 16 October 1937) was Head Brewer of Guinness, Head Experimental Brewer of Guinness, and father of modern British statistics. He pioneered small sample experimental design and analysis with an economic approach to the logic of uncertainty. Gosset published under the pen name “Student,” and developed most famously Student’s t-distribution - originally called Student’s “z” - and “Student’s test of statistical significance”” (Wikipedia)

Distribuția T

- cozi (tails) mai pronunțate decât distribuția Z
- spre deosebire de distribuția normală, caracterizată de 2 parametri (μ, σ^2), vom discuta de distribuția T ca fiind centrată în jurul lui 0, cu o formulă specifică pentru scală
- caracterizată de 1 parametru, **degrees of freedom** - df
- pe măsură ce df crește, distribuția se comportă din ce în ce mai mult ca o distribuție normală standard

Distribuția T (2)

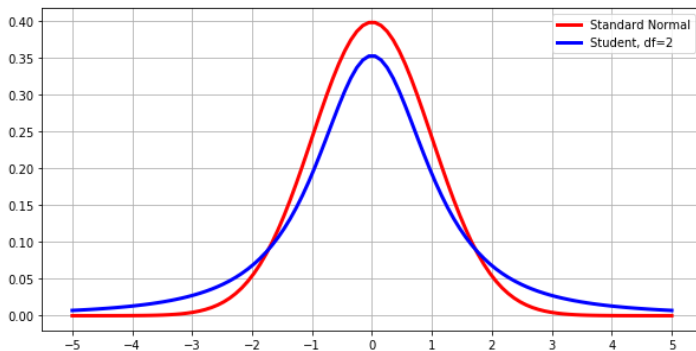
- distribuția sample mean normalizată:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

- folosind estimandul S în loc de σ , această distribuție nu e standard normală (ar fi normală folosind σ), ci
- urmează o distribuție T Gosset cu $n - 1$ grade de libertate
- intervalul de confidență T:

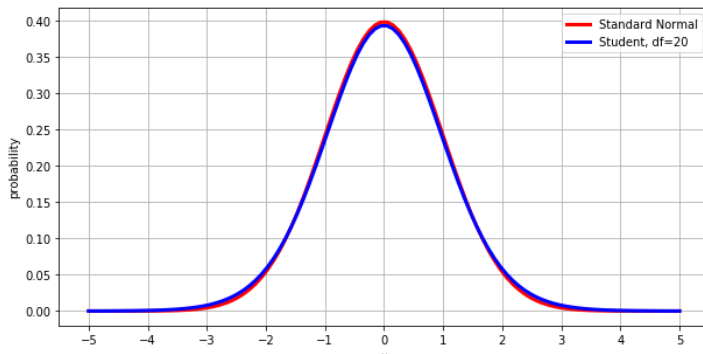
$$\bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}}, \text{ unde } t_{n-1} \text{ este quantila relevantă}$$

Distribuția T (3)



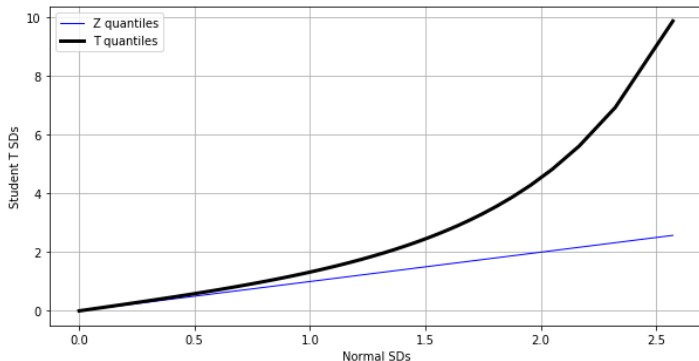
- distribuția T cu $df=2$

Distribuția T (4)



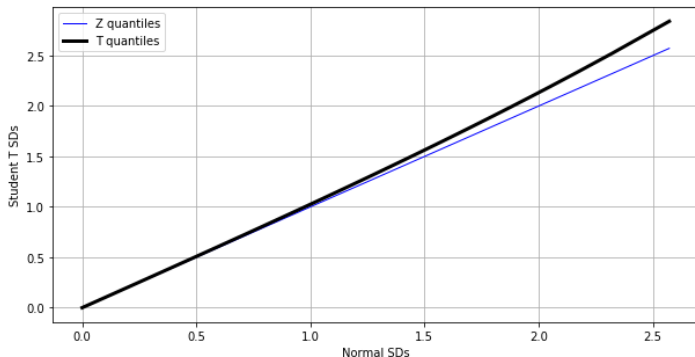
- distribuția T cu $df=20$

Quantilele distribuției normale standard vs. cele ale distribuției Student T



- distribuția Student T cu 2 grade de libertate
- doar două grade de libertate înseamnă doar 3 puncte în sample pentru estimarea dispersiei S !

Quantilele distribuției normale standard vs. cele ale distribuției Student T (2)



- distribuția Student T cu 20 de grade de libertate
- întotdeauna T quantile va fi peste Z quantile
- intervalele T vor fi **mai largi** decât cele Z
- distribuția T este mai incertă decât distribuția Z

Intervalul T

- intervalul T presupune că datele sunt iid
- funcționează bine când distribuția datelor este (aproape) simetrică și sub forma unei 'cocoase de cămilă'
- perechile de observații sunt analizate folosind intervalul T prin calcularea diferențelor față de observațiile inițiale (sau diferențele între log-valori)
- pentru valori mari pentru df, quantilele T devin foarte apropiate de quantilele Z, așadar intervalele de confidență T converg la același interval descris de CLT
- pentru distribuții asimetrice, nu are sens centrarea intervalului la medie
 - în acest caz putem încerca folosind log-valori sau folosind o altă procedură (ex. bootstrap¹ confidence intervals)

¹ "resample the data with replacement, and the size of the resample must be equal to the size of the original data set. Then the statistic of interest is computed from the resample from the first step" (Wikipedia)

Sleep dataset

- sleep.csv, setul de date folosit de W. Galton pentru analizarea creșterii numărului de ore de somn pentru 10 pacienți care au luat două medicamente, care acționează ca somnifere

	Unnamed: 0	extra	group	ID
0	1	0.7	1	1
1	2	-1.6	1	2
2	3	-0.2	1	3
3	4	-1.2	1	4
4	5	-0.1	1	5

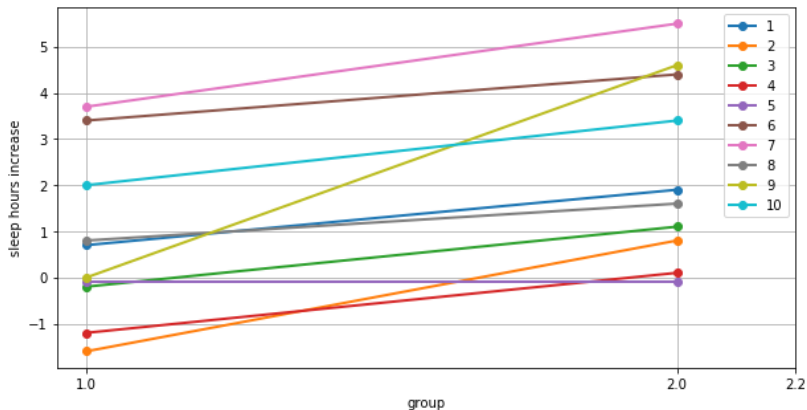
	Unnamed: 0	extra	group	ID
10	11	1.9	2	1
11	12	0.8	2	2
12	13	1.1	2	3
13	14	0.1	2	4
14	15	-0.1	2	5

Sleep dataset - join

```
sleep[sleep.group==1]\
  .set_index('ID')\
  .join(
    sleep[sleep.group==2].set_index('ID'),
    lsuffix='_before',
    rsuffix='_after'
  )
```

	Unnamed: 0_before	extra_before	group_before	Unnamed: 0_after	extra_after	group_after
ID						
1	1	0.7	1	11	1.9	2
2	2	-1.6	1	12	0.8	2
3	3	-0.2	1	13	1.1	2
4	4	-1.2	1	14	0.1	2
5	5	-0.1	1	15	-0.1	2
6	6	3.4	1	16	4.4	2
7	7	3.7	1	17	5.5	2
8	8	0.8	1	18	1.6	2
9	9	0.0	1	19	4.6	2
10	10	2.0	1	20	3.4	2

Sleep dataset: paired T test



- se pot compara grupurile fie din punctul de vedere al dispersiei, fie al dispersiei diferențelor (mult mai mică, observațiile sunt corelate)

Sleep dataset: 95% T confidence interval

```
1 diff = y2 - y1
2 m = np.mean(diff); s = np.std(diff)
3 print(m + np.array([-1, 1]) * stats.t.ppf(0.975, df=9) * s / np.sqrt(10))
4 print(stats.t.interval(0.95, loc=m, scale=s/np.sqrt(10), df=9))
```

```
[0.74526707 2.41473293]
(0.7452670722078104, 2.4147329277921896)
```

- dacă am face resampling din grupuri, în 95% din cazuri am fi siguri că media diferențelor este în acest interval
- 0 nu face parte din interval, medicamentul prelungește somnul

A/B testing

- presupunem că dorim să comparăm media BP între două grupuri de pacienți într-un trial randomizat
- cei care au primit tratamentul și cei care au primit placebo
- cele două grupuri sunt randomizate, ne așteptăm ca media să aibă distribuție T
- nu putem folosi paired T test pentru că nu există potrivire între grupuri

A/B testing: T confidence intervals pentru grupuri independente

- intervalul de confidență pentru grupuri independente presupune calcularea standard error a diferenței:

$$\bar{Y} - \bar{X} \pm t_{n_x+n_y-2, 1-\alpha/2} S_p \left(\frac{1}{n_x} + \frac{1}{n_y} \right)^{1/2}$$

- S_p este denumită **pooled standard deviation**, iar S_p^2 - **pooled variance**

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

- se face presupunerea că **dispersia în cele două grupuri este aceeași**, deci pooled variance este o medie a dispersiilor celor două grupuri
- n_x, n_y ponderează această medie cu dimensiunea sample-urilor

Exemplu: oral contraceptive și BP

- A/B testing pe seturi randomizate de pacienți, să determinăm dacă un contraceptiv oral are ca efect creșterea sau nu a tensiunii arteriale (BP)
- două seturi, OC - oral contraceptive și C - control

```

1 # independent group A/B testing
2 x_oc, s_oc = 132.86, 15.34
3 x_c, s_c = 127.44, 18.23
4 n_oc, n_c = 8, 21
5 sp = np.sqrt( ((n_oc-1)*s_oc**2 + (n_c-1)*s_c**2) / (n_oc+n_c-2) )
6 x_oc - x_c + np.array([-1, 1]) * stats.t.ppf(.975, df=n_oc+n_c-2) \
7     * sp * (1/n_oc+1/n_c)**.5

```

array([-9.52109748, 20.36109748])

- intervalul conține 0, nu se poate exclude posibilitatea ca medicamentul să nu aibă nici o influență asupra BP

Exemplu: sleep data ca test T independent

```

1 # sleep data - independent si paired
2 y1 = sleep1['extra_before'].values
3 y2 = sleep1['extra_after'].values
4
5 x_1, s_1 = np.mean(y1), np.std(y1)
6 x_2, s_2 = np.mean(y2), np.std(y2)
7 n = len(y1)
8 sp = np.sqrt( (s_1**2 + s_2**2) / 2 )
9 print(x_2 - x_1 + np.array([-1, 1]) * stats.t.ppf(.975, df=2*n-2) \
10       * sp * np.sqrt(2/n))
11 print(stats.t.interval(0.95, loc=m, scale=s/np.sqrt(10), df=9))

```

[-0.1123315 3.2723315]
 (0.7452670722078104, 2.4147329277921896)

- dacă facem abstracție de pairing și presupunem grupurile independente (ceea ce nu e cazul), intervalul conține valoarea 0 (somniferul nu are efect)
- pe figură, variabilitatea e explicată de variabilitatea dintre seturi

A/B testing: independent groups, unequal variances

- media populației:

$$\bar{Y} - \bar{X} \pm t_{df} \left(\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y} \right)^{1/2}$$

- standard error este calculată presupunând că cele două grupuri au standard error diferite
- problema e că în acest caz, statistica normalizată nu urmează o distribuție T; ea poate fi aproximată de o distribuție T în care gradele de libertate se calculează ca:

$$df = \frac{(S_x^2/n_x + S_y^2/n_y)^2}{\left(\frac{S_x^2}{n_x}\right)^2 / (n_x - 1) + \left(\frac{S_y^2}{n_y}\right)^2 / (n_y - 1)}$$

Exemplu: unequal variances

```

1 # oral contraceptive example - independent test, unequal variances
2 x_oc, s_oc = 132.86, 15.34
3 x_c, s_c = 127.44, 18.23
4 n_oc, n_c = 8, 21
5 df = (s_oc**2/n_oc + s_c**2/n_c)**2 / \
6       ( (s_oc**2/n_oc)**2/(n_oc-1) + (s_c**2/n_c)**2/(n_c-1) )
7 print('df: ', df)
8 sp = np.sqrt( s_oc**2/n_oc + s_c**2/n_c )
9 print(x_oc - x_c + np.array([-1, 1]) * stats.t.ppf(.975, df=df) * sp)
10 print(stats.t.interval(
11     0.95,
12     loc = x_oc - x_c,
13     scale = np.sqrt(s_oc**2/n_oc + s_c**2/n_c),
14     df=df
15 ))

```

```

df: 15.035180011917006
[-8.91332743 19.75332743]
(-8.913327425893712, 19.753327425893744)

```

- dacă dispersiile nu sunt egale, intervalul de confidență obținut e mai îngust, dar tot nu exclude 0

Intervale de confidență T: utilizări

- T interval pentru: observații singulare sau observații paired (folosim diferența acestor observații)
- ca variante, folosim pooled variance (observații independente, aceeași dispersie în cele două grupuri) și grade de libertate specifice (dispersie diferită)
- skewed data: log-valori, bootstrap CI

Compararea altor tipuri de date

- pentru variabile binomiale, există multe posibilități de a compara două grupuri:
 - rata de risc, diferența de risc, odds ratio
 - Chi-squared tests
- pentru rate, folosim distribuții Poisson

1 T confidence intervals

2 Hypothesis testing

3 P-values

4 Bibliografie

Hypothesis testing

- luarea de decizii folosind datele²
- status quo-ul reprezintă null hypothesis, notată cu H_0
- ipoteza nulă e presupusă adevărată și se caută dovezi statistice pentru a o rejecta în favoarea unei ipoteze alternative

²ML much? :)

Exemplu: hypothesis testing

- numărul de anomalii respiratorii, o rată de mai mult de 30 de evenimente pe oră, e considerat un factor serios de tulburare a somnului
- presupunem că într-un sample de 100 de pacienți supraponderali cu alți factor de risc pentru tulburări ale somnului, numărul mediu de astfel de incidente respiratorii este 32 / oră, cu o deviație standard de 10 evenimente / oră
- construim ipoteza că:
 - $H_0 : \mu = 30$ (ipoteza nulă)
 - $H_a : \mu > 30$ (ipoteza alternativă)
 - unde μ este media populației care are aceste tulburări de respirație

Hypothesis testing (2)

- ipoteza alternativă poate fi $<$, $>$ sau \neq
- există patru posibilități ale experimentului nostru statistic:

Adevărat	Decizie	Rezultat
H_0	H_0	ipoteză nulă corect acceptată
H_0	H_a	Type I error
H_a	H_a	ipoteză nulă corect rejectată
H_a	H_0	Type II error

- vom controla probabilitatea Type I error să fie mică
- cele două tipuri de erori sunt legate: când Type I error rate crește, Type II error rate scade și invers

Exemplu - tribunal

- tribunal; ipoteza nulă este că acuzatul e nevinovat
- impunem un standard asupra dovezilor pentru a rejecta ipoteza nulă
- standard setat prea jos:
 - crește procentul inocenților condamnați (type I error)
 - crește procentul vinovaților condamnați (rejectarea corectă a H_0)
- standard setat prea sus:
 - crește procentul inocenților lăsați liberi (acceptarea corectă a H_0)
 - crește procentul vinovaților lăsați liberi (type II error)
- cum avem un standard mai bun ? creștem mărimea sample-ului

Exemplu: rata de evenimente respiratorii

- un standard bun pentru rejectarea ipotezei nule ar fi ca media ratei, \hat{X} , să fie mai mare ca o constantă C , care ia în considerare variabilitatea lui \bar{X}
- de obicei se ia C astfel încât probabilitatea type I error, α este un număr mic (5%): $P(H_a | H_0 = \text{true}) < 0.05$
- $\alpha = \text{type I error rate} = \text{probabilitatea de a rejecta ipoteza nulă } H_0 \text{ când aceasta e adevărată}$

Exemplu (2)

- pentru distribuția ratei evenimentelor respiratorii în populație, eroarea standard a mediei este $10/\sqrt{100} = 1$
- H_0 este că $X \sim N(30, 1)$
- dorim să alegem C astfel ca $P(\bar{X} > C; H_0) = 5\%$
- percentila 95% a distribuției standard normale este 1.645 deviații standard în dreapta mediei
- $C = 30 + 1 * 1.645 = 31.645$
- probabilitatea ca $\bar{X} \sim N(30, 1) > C$ este 5%
- regula “rejectăm H_0 dacă $\bar{X} \geq 31.645$ ” are proprietatea că vom rejecta H_0 5% din timp dacă H_0 este de fapt adevărată

Exemplu (3)

- rejectarea se face pe baza Z-score: calculăm cu câte deviații standard media sample-ului este mai departe de media populației

$$\frac{32 - 30}{10/\sqrt{100}} = 2$$

- este mai mare ca 1.645; adică:

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > Z_{1-\alpha}$$

- șansa ca această inegalitate să se producă considerând ipoteza H_0 adevărată este mai mică decât 5%

Exemplu revizitat

- pentru același exemplu, consierăm doar 16 pacienți (față de 100 anterior)
- statistica:

$$\frac{\bar{X} - 30}{s/\sqrt{16}}$$

- , urmează o distribuție T cu 15 grade de libertate sub ipoteza nulă H_0
- sub ipoteza H_0 , probabilitatea ca statistica să fie mai mare ca percentila 95% a distribuției T este 5%
- percentila 95% a distribuției T cu 15 grade de libertate este 1.7531
- statistica este $(32 - 30)/10/\sqrt{16} = 0.8$
- deoarece $0.8 < 1.7531$, eșuăm în a rejecta H_0 (fail to reject)

Two-sided test

- uneori este necesar să testăm nu numai dacă $\mu > 30$, ci și că $\mu \neq 30$, adică dacă media este prea mare sau prea mică
- ipoteza alternativă este $H_a : \mu \neq 30$
- probabilitatea de rejectare în cazul H_0 trebuie să fie 5%, împărțită egal, 2.5% pentru upper tail și 2.5% pentru lower tail
- astfel, rejectăm dacă statistica determinată este mai mare decât quantila 0.975 sau mai mică decât quantila 0.025
- același fel cu a spune că valoarea absolută a statisticii este mai mare decât quantila 0.975, adică 1.96
- dacă eșuăm în a rejecta one-sided test, vom eșua în a rejecta two-sided test (acesta din urmă e și mai strict)

Exemplu fathers and sons

```

1  # calculăm diferența dintre înălțimi
2  fs = father_son['sheight'].values - father_son['fheight'].values
3
4  # statistica,  $(X - \text{mean\_pop}) / \text{std\_sample} / \sqrt{\text{sample length}}$ 
5  # atenție la calculul standard error, se face pe n-1, unde
6  # n este dimensiunea sample-ului
7  print((np.mean(fs) - 0) * np.sqrt(len(fs)) / np.std(fs, ddof=1))
8
9  # folosim paired (related) T test pentru a calcula statistica
10 # ipoteza  $H_0$  este ca mediile celor doua populatii sa nu difere
11 print(stats.ttest_rel(
12     father_son['sheight'].values,
13     father_son['fheight'].values\
14 ).statistic)
15
16 # alternativ avem si one sample T test, specificam media populatiei
17 # ipoteza  $H_0$  este ca media diferentelor sa fie 0
18 print(stats.ttest_1samp(
19     father_son['sheight'].values - father_son['fheight'].values,
20     popmean=0
21 ).statistic)

```

11.788549038237015

11.788549038237013

11.788549038237013

Exemplu fathers and sons (2)

```

1  # pentru ipoteza  $H_0$ , dacă ea e adevărată (diferența e zero),
2  # în 95% din cazuri media va fi sub 1.6463
3  print(stats.t.ppf(0.95, df=len(fs)-1))
4
5  # putem calcula și intervalul T de confidență 95% pentru media populației:
6  print(stats.t.interval(
7      0.95,
8      loc = np.mean(fs),
9      scale = np.std(fs, ddof=1)/np.sqrt(len(fs)),
10     df=len(fs)-1
11 ))
12
13 # se vede că:
14 # 1) statistica 11.78 > 1.64, rejectăm  $H_0$ 
15 # 2) zero nu intră în acest interval
16 # 3) df = 1077, distribuția e aproape normală

```

1.6462696795616023
(0.831029622612876, 1.1629159618027087)

- intervalul de confidență e o verificare pentru rejectarea/acceptarea ipotezei H_0

Exemplu revizitat: rata de evenimente respiratorii

```

1 # rata de evenimente respiratorii revizitata
2 print('statistic: ', (32 - 30)*np.sqrt(16)/10)
3 print('95% percentile:', stats.t.ppf(0.95, df=15))
4 print('reject H_0'
5       if (32 - 30)*np.sqrt(16)/10 > stats.t.ppf(0.95, df=15)
6       else 'fail to reject H_0'
7 )
8 print('verificare: media poate fi in intervalul\n\t',
9       stats.t.interval(
10          0.95,
11          loc = 30,
12          scale = 32/np.sqrt(16),
13          df=15
14      )
15 )

```

```

statistic: 0.8
95% percentile: 1.7530503556925547
fail to reject H_0
verificare: media poate fi in intervalul
            (12.948403635525416, 47.051596364474584)

```

Legătura cu intervalul de confidență

- considerăm ipoteza $H_0 : \mu = \mu_0$ vs. $H_a : \mu \neq \mu_0$
- dacă luăm setul tuturor valorilor posibile pentru care eșuăm să rejectăm H_0 , acest set este $(1 - \alpha)100\%$ intervalul de confidență³ pentru μ
- reciproc, dacă intervalul $(1 - \alpha)100\%$ conține μ_0 , atunci eșuăm să rejectăm H_0

³de regulă $\alpha = 0.05$

Two group intervals

- știm să realizăm two group T tests, pentru că am acoperit realizarea intervalelor de confidență T pentru grupuri paired, independente cu aceeași dispersie respectiv independente cu dispersii diferite
- regulile de rejectare sunt aceleași
- ipoteza null este acum $H_0 : \mu_1 = \mu_2$
- statistica testului este:

$$\frac{\bar{X}_1 - \bar{X}_2 - 0}{\text{standard error of the mean}}$$

- media estimată este acum $\bar{X}_1 - \bar{X}_2$
- media populației este 0

Exemplu revizitat: oral contraceptive

```

1 # oral contraceptive example - independent test, unequal variances
2 x_oc, s_oc, n_oc = 132.86, 15.34, 8
3 x_c, s_c, n_c = 127.44, 18.23, 21
4 df = (s_oc**2/n_oc + s_c**2/n_c)**2 / \
5       ((s_oc**2/n_oc)**2/(n_oc-1) + (s_c**2/n_c)**2/(n_c-1) )
6 sp = np.sqrt( s_oc**2/n_oc + s_c**2/n_c )
7
8 statistic = stats.ttest_ind_from_stats(
9     x_oc, s_oc, n_oc, x_c, s_c, n_c, equal_var=False).statistic
10 print('df: ', df); print('statistic: ', statistic)
11 percentile = stats.t.ppf(0.95, df=df)
12 print('95% percentile:', percentile)
13 print('reject H_0' if statistic > percentile else 'fail to reject H_0')
14 print('T confidence interval: ', stats.t.interval(
15     0.95, loc = x_oc - x_c, scale = np.sqrt(s_oc**2/n_oc + s_c**2/n_c), df=df
16 ))
17 print('two group population mean: ', 0)

```

```

df: 15.035180011917006
statistic: 0.805821448286547
95% percentile: 1.7527810366681962
fail to reject H_0
T confidence interval: (-8.913327425893712, 19.753327425893744)
two group population mean: 0

```

Test binomial

Reluăm problema cu acel prieten care are 8 copii, din care 7 sunt fete și nici unele gemene.

Realizați testarea ipotezei $H_0 : p = 0.5$, $H_a : p > 0.5$

Care este regiunea de rejectare relevantă astfel ca probabilitatea de rejectare este mai mică decât 5% ?

Căutăm probabilitatea, considerând H_0 adevărată, ca cuplul să aibă 7 sau mai multe fete.

```
1 print(pd.DataFrame(list(zip(
2     ['[%d : 8]' % i for i in range(0, 9)],
3     [round(1 - stats.binom.cdf(i-1, n=8, p=0.5), 4) for i in range(0, 9)]
4 )), columns=['Rejection Region', 'Type I Error Rate']))
```

	Rejection Region	Type I Error Rate
0	[0 : 8]	1.0000
1	[1 : 8]	0.9961
2	[2 : 8]	0.9648
3	[3 : 8]	0.8555
4	[4 : 8]	0.6367
5	[5 : 8]	0.3633
6	[6 : 8]	0.1445
7	[7 : 8]	0.0352
8	[8 : 8]	0.0039

Test binomial (2)

- e imposibil să atingem nivelul testului de exact 5% datorită discretizării
- cea mai apropiată regiune de rejecție este $[7 : 8]$
- orice α sub 0.0039 nu este posibil
- pentru sample-uri de mărimi mai mari, putem face o aproximare cu o distribuție normală (vezi exemplul cu zarurile)
- two sided test nu e evident; pentru un astfel de test, putem lua valorile probabilităților p_0 pentru care eșuăm să rejectăm, pentru a obține un confidence interval binomial exact
- pentru acest lucru se folosește P-value

- 1 T confidence intervals
- 2 Hypothesis testing
- 3 P-values**
- 4 Bibliografie

P-values

- este cea mai comună măsură a semnificației statistice
- sunt peste tot; sunt controversate ca folosire, privind interpretarea și folosința
- presupunerea de la care pleacă: presupunem că nu se întâmplă nimic (H_0) - cât de neobișnuit este să observăm estimarea pe care am obținut-o?
 - definim distribuția ipotetică a datelor (statistic) când “nimic special nu se întâmplă” - H_0
 - calculăm variabilele statistice cu datele pe care le avem (T test statistic)
 - comparăm ce am calculat cu distribuția ipotetică și vedem dacă valoarea este extremă - P-value - probabilitatea de a obține o valoare ca aceasta sau mai extremă⁴
 - P-value este probabilitatea de a observa o valoare ca cea obținută sau mai extremă; este mică dacă H_0 este adevărată

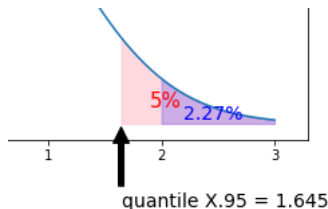
⁴mai spre coadă / tail

P-values (2)

- P-value este probabilitatea, dacă H_0 e adevărată, de a obține o statistică T test la fel de extremă sau mai extremă decât cea obținută
- dacă P-value este mică, atunci fie H_0 este adevărată și am observat un eveniment rar, sau H_0 este falsă
- exemplu:
 - presupunem că obținem o statistică T de 2.5 pentru 15 grade de libertate, testând $H_0 : \mu = \mu_0$ vs. $H_a: \mu > \mu_0$
 - care este probabilitatea de a obține o statistică T mai mare ca 2.5 ?
 - valoarea e dată de survival function, $1 - \text{stats.t.cdf}(2.5, \text{df}=15)$ adică 0.01225
 - deci probabilitatea de a obține o valoare ca asta sau mai extremă este de 1%
 - atunci: fie H_0 este adevărată și am observat o statistică T excepțional de mare (de extremă), fie H_0 este falsă

Nivelul de signifiacance atins

- să zicem că statistica a fost 2 pentru $H_0 : \mu_0 = 30$ vs. $H_a : \mu > 30$
- atunci rejectăm H_0 , testul one sided pentru $\alpha = 0.05$, vom rejecta și pentru $\alpha = 0.04$, ș.a.m.d. până la $\alpha = 0.0227$



- dacă mergem cu α mai în dreapta lui 0.0227, eșuăm în a rejecta H_0
- pentru valori mai mici ca P-value rejectăm, peste P-value - nu

- probabilitatea de 2.27% este probabilitatea de a întâlni o statistică T egală cu 2 sau mai extremă
- un alt mod de gândire: definim P-value ca fiind cea mai mică valoare pentru α pentru care am rejecta $H_0 =$ “attained significance level”
- P-value este o statistică a testului T foarte convenabil de comunicat altora, deoarece ei o pot testa vs. orice valoare α pe care o doresc

Comentariu

- prin raportarea P-value, se poate realiza testarea ipotezei la orice α dorit, nu neapărat 5%
- **regula**: dacă P-value este mai mic decât α , se rejectează H_0 , altfel “fail to reject”
- pentru two sided test, ne interesează probabilitatea ca z să fie mai mare sau egală cu 2 respectiv mai mică sau egală cu -2, deci din această cauză valoarea P-value **se dublează** (adunăm ariile din cozi)
- majoritatea software-ului lucrează implicit cu two-sided test P-values

Exemplu revizitat: prietenul cu 7 fete din 8 copii

- presupunem că un prieten are 8 copii, din care 7 fete, fără gemeni
- dacă fiecare gen are 50% probabilitate la naștere, care este probabilitatea de a avea 7 sau mai multe fete din 8 copii?
- testăm ipoteza $H_0 : p = 0.5$ vs. $H_a : p > 0.5$, adică vrem să obținem care este probabilitatea de a observa 7 sau 8 fete din totalul de 8 copii?

```
1 # 7 fete din 8 revisited
2 print(1 - stats.binom.cdf(6, n=8, p=0.5))
3 print(stats.binom_test(7, n=8, p=0.5, alternative='greater'))
```

```
0.03515625
```

```
0.03515625
```

- rejectăm dacă type I error rate este 5%, dar nu și, de exemplu, pentru type I error rate de 1% (vrem să fim 99% siguri că probabilitatea NU este fair, 0.5)
- two sided value (P-value) este dublul valorii celei mai mici pentru probabilitățile de a fi mai mică sau mai mare ca 7 (vezi statinf4.ipynb)

Exemplu Poisson

- un spital are o rată de infectări de 10 infecții la 100 persoane (rata de 0.1) pe parcursul ultimei perioade de monitorizare
- rata de 0.05 este un benchmark important
- nu putem lua această decizie pe baza fluctuațiilor aleatoare
- testarea ipotezei ia în considerare incertitudinea
- rata infecțiilor este distribuită Poisson
- $H_0 : \lambda = 0.05$ vs. $H_a : \lambda > 0.05$
- considerând această rată de 0.05 (la limită), care e probabilitatea de a obține 10 sau mai multe infecții; este ea mai mare decât 5%?

```

1 # infection rate
2 # distribuțiile discrete ne dau  $P(X \leq x)$ , de aceea luăm de la 9
3 # avem nevoie de survival function pentru ca e upper tail
4 1 - stats.poisson.cdf(9, mu=0.05 * 100)

```

0.03182805730620486

- P-value destul de mic, e destul de puțin probabil să fi observat 10 infecții dacă rata acceptată este de 5 infecții
- reject H_0 , trebuie sesizată conducerea spitalului

- 1 T confidence intervals
- 2 Hypothesis testing
- 3 P-values
- 4 Bibliografie

Bibliografie

- `https://en.wikipedia.org/wiki/Bootstrapping_(statistics)`
- `https://en.wikipedia.org/wiki/Risk_ratio`
- `https://en.wikipedia.org/wiki/Risk_difference`
- `https://en.wikipedia.org/wiki/Odds_ratio`