

برآورد نقطه ای میانگین و توزیع آن



- فرض کنید جامعه ای دارای سه عضو باشد $N=3$ و از آنها درباره تعداد نخ سیگارهایی که در یک شبانه روز می کشند سوال کرده ایم.
 $\{7, 25, 37\}$

$$\mu = \frac{7 + 25 + 37}{3} = 23$$

$$\sigma^2 = \frac{1}{3} \{ (7 - 23)^2 + (25 - 23)^2 + (37 - 23)^2 \} = 152$$

در نمونه گیری با جایگذاری 9 حالت ممکن است پیش آید.

	7	25	37
7	(7,7)	(7,25)	(7,37)
25	(25,7)	(25,25)	(25,37)
37	(37,7)	(37,25)	(37,37)

میانگین هر حالت در جدول زیر نشان داده شده است

	7	25	37
7	7	16	22
25	16	25	31
37	22	31	37

در جدول زیر تابع توزیع احتمال برای پارامتر میانگین جامعه نشان داده شده است.

$\bar{X}$	7	16	22	25	31	37
$P(\bar{X} = \bar{x})$	1/9	2/9	2/9	1/9	2/9	1/9

برای به دست آوردن میانگین میانگین ها از فرمول زیر و به کمک تابع توزیع احتمال می توان استفاده نمود.



$$E(\bar{X}) = \sum p_i \bar{x}_i = \frac{1}{9} \times 7 + \frac{2}{9} \times 16 + \dots + \frac{1}{9} \times 37 = 23$$

$$\begin{aligned}\sigma^2(\bar{X}) &= \sum p_i (\bar{X}_i - E(\bar{X}_i))^2 \\ &= \frac{1}{9} \times (7 - 23)^2 + \frac{2}{9} \times (16 - 23)^2 + \dots + \frac{1}{9} \times (37 - 23)^2 \\ &= 76\end{aligned}$$

نتیجه به دست آمده



$$E(\bar{X}) = E(X) = 23$$

$$\sigma^2(\bar{X}) = \frac{\sigma^2}{n} = \frac{152}{2} = 76$$

قضیه حد مرکزی



- اگر متغیر تصادفی X دارای توزیعی با میانگین μ و انحراف معیار σ باشد. میانگین توزیع نمونه ای $\bar{X}$ که بر اساس نمونه ای به حجم n بدست می آید دارای توزیعی با میانگین μ و انحراف $\frac{\sigma}{\sqrt{n}}$ خواهد بود که با بزرگ شدن n به سمت توزیع نرمال میل می کند.

مثال



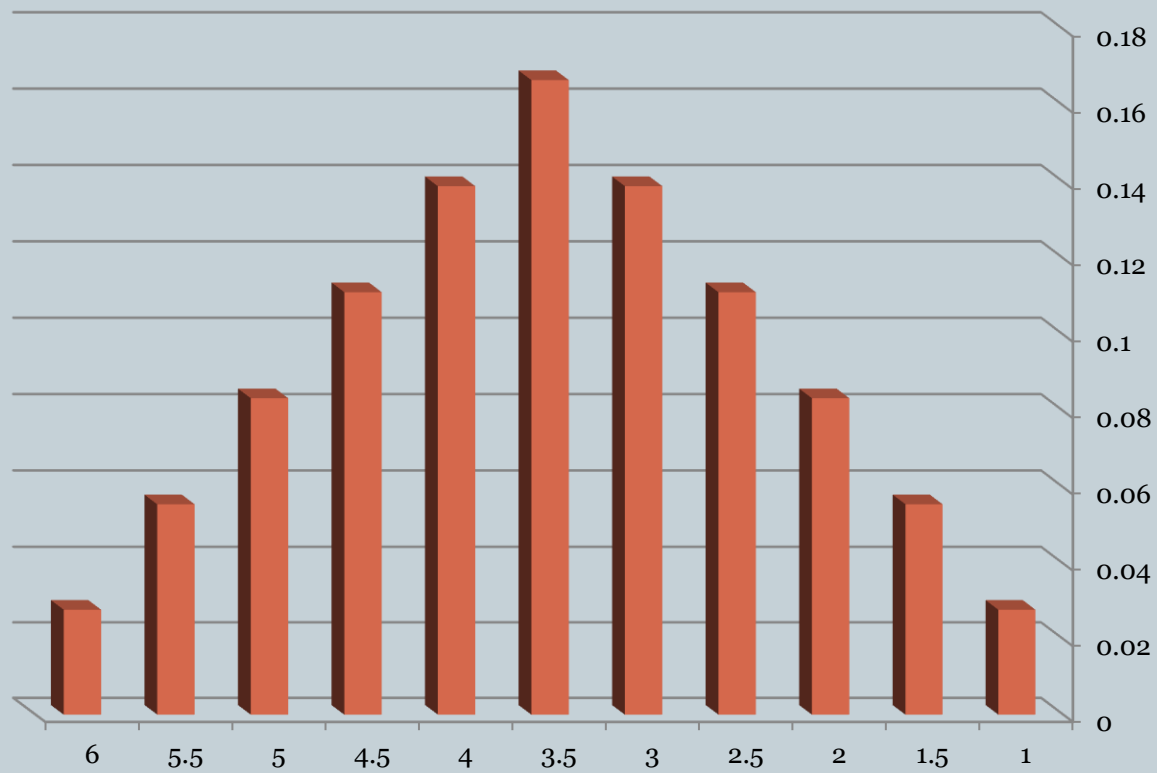
• در پرتاب دو تاس فضای نمونه

$$S = \{(1,1), (1,2), (1,3), \dots, (1,6), \dots, (6,6)\}$$

$X =$ میانگین اعداد دو تاس

$$X = 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6$$

$X =$	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
$P(X=x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36



اندازه نمونه چقدر باشد تا توزیع نرمال گردد ؟



- اگر توزیع متغیر نرمال باشد ، میانگین هر تعداد نمونه از توزیع نرمال دارای توزیع نرمال است.
- معمولاً برای اندازه نمونه های بزرگتر از 30 میانگین نمونه دارای توزیعی نرمال است.

برآورد فاصله‌ای میانگین (فاصله اطمینان)

در برآورد فاصله‌ای برای میانگین هدف ارائه یک بازه می‌باشد. بطوریکه چه مقدار به میانگین نمونه اضافه و کم کنیم تا با احتمال $1 - \alpha$ (خطای نوع اول است) $\bar{x}$ در این بازه قرار گیرد. جهت برآورد فاصله برای میانگین جامعه بسته به اینکه واریانس جامعه معلوم باشد و یا معلوم نباشد توزیع $\bar{x}$ به ترتیب نرمال یا در حالت دوم توزیع t می‌باشد.

در حالت اول و قتیکه واریانس جامعه معلوم باشد برآورد فاصله‌ای میانگین با ضریب اطمینان $1 - \alpha$ به صورت زیر بدست می‌آید :

$$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

که در این فرمول $\bar{x}$ میانگین صفت مورد نظر در نمونه انتخابی، σ انحراف معیار جامعه، n اندازه نمونه و $Z_{1-\frac{\alpha}{2}}$ عبارتست از مقداری از کمیت که سطح زیر منحنی نرمال استاندارد از $-\infty$ تا این نقطه برابر است با $1 - \frac{\alpha}{2}$ می‌باشد.

مثال : صفت x در جامعه داراي توزيع نرمال با انحراف معيار $\sigma = 10.5$ مي باشد.
از اين جامعه يك نمونه تصادفي 25 تايي انتخاب کرده ايم و مشاهده مي شود که $\bar{x} = 40$ است. مطلوب است تعيين حدودي که با 95/0 اطمینان میانگین جامعه را در بر گیرد.

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$$

$$Z_{1-\frac{\alpha}{2}} = Z_{1-0.025} = Z_{0.975} = 1.96$$

$$40 \pm 1.96 \left(\frac{10.5}{\sqrt{25} = 5} \right) = (35.884, 44.116)$$

برآورد فاصله‌ای برای میانگین (واریانس جامعه نامعلوم)



- در این حالت جهت ساخت فاصله اطمینان برای میانگین از رابطه زیر استفاده می‌کنیم :

$$\bar{x} \pm t_{\left(1-\frac{\alpha}{2}, n-1\right)} \frac{S}{\sqrt{n}}$$

- بطوریکه در اینجا S برآورد نقطه‌ای انحراف معیار یعنی انحراف معیار نمونه می‌باشد و t عبارت است از مقداری از کمیت سطح زیر منحنی توزیع t از $-\infty$ تا این نقطه که برابر با $1-\frac{\alpha}{2}$ می‌باشد.
- $n-1$ که در اندیس t آمده است پارامتری بنام درجه آزادی برای توزیع t می‌باشد. توجه داشته باشیم که هرگاه حجم نمونه بیشتر از 30 باشد توزیع t به توزیع Z یعنی نرمال استاندارد نزدیک می‌شود و در این صورت جهت برآورد فاصله اطمینان برای میانگین می‌توان همانند قبل برای وقتی که واریانس معلوم باشد استفاده کرد.

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
Z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%

مثال : فرض کنید فشار خون دیاستولیک مردان در جامعه‌ای از توزیع نرمال با واریانس $\sigma = 10.4$ پیروی می‌کند. در یک نمونه 25 تایی متوسط فشار خون افراد انتخابی برابر با 8.5 شده است. مطلوب است برآورد فاصله‌ای 90% برای میانگین جامعه.

$$\bar{x} \pm Z_{1-\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}} \right)$$

$$1 - \alpha = 0.90 \Rightarrow \alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05$$

$$Z_{1-\frac{\alpha}{2}} = Z_{0.95} = 1.64$$

$$8.5 \pm 1.64 \left(\frac{\sqrt{10.4}}{\sqrt{25}} \right) =$$

$$8.5 \pm 1.64(0.64) =$$

$$(7.44, 9.56)$$

مثال : مدت تسکین درد توسط یک داروی جدید برای 6 بیمار مبتلا به رماتیسم مفصلي بر حسب ساعت به قرار زیر بوده است : $4/2$, $9/4$, $5/2$, $7/3$, $3/4$, $2/2$,
مطلوب است يك فاصله اطمینان 99% برای میانگین

$$\bar{x} = \frac{20}{6} = 3.33$$

$$S^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{(n-1)} = \frac{(73.4) - \frac{(400)^2}{6}}{(5)} = 1.27 \Rightarrow S = 1.13$$

$$1 - \alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005$$

$$t_{\left(1-\frac{\alpha}{2}, n-1\right)} = t_{(0.995, 5)} = 4.03$$

$$\bar{x} \pm t_{\left(1-\frac{\alpha}{2}, n-1\right)} \frac{S}{\sqrt{n}}$$

$$3.33 \pm 4.03 \left(\frac{1.13}{\sqrt{6}} \right) =$$

$$3.33 \pm 4.03(0.46) =$$

$$(3.14, 5.19)$$

برآورد فاصله اطمینان تفاضل میانگین دو جامعه وابسته (زوجی)



- در این حالت اختلاف بین دو گروه به دست آورده و مانند یک نمونه برای اختلاف ها فاصله اطمینان به دست می آوریم.
- در مطالعه ای به منظور کاهش فشار خون از یک دوره ورزش آیربیک استفاده شده است و فشار خون افراد قبل از شرکت در کلاس های ورزشی و یک ماه پس از شرکت در کلاس های ورزشی در جدول زیر نشان داده شده است یک فاصله اطمینان 99 درصدی برای متوسط اختلاف فشار خون به دست آورید.

قبل	12.5	11.3	13.6	14	15.2	14.6	13.8	15	14.7	13.5	12.2	13.7
بعد	11.5	10.8	13	14.5	14	13	13.5	13.2	13.5	15	11.5	14
diff	1	0.5	0.6	-0.5	1.2	1.6	0.3	1.8	1.2	-1.5	0.7	-0.3

$$\sum d_i = 6.6 \quad S_d^2 = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1} = \frac{13.46 - \frac{6.6^2}{12}}{11} = 0.89$$

$$\sum d_i^2 = 13.46 \quad \bar{d} \pm t_{\left(1-\frac{\alpha}{2}, n-1\right)} \frac{S_d}{\sqrt{n}} = 0.55 \pm t_{(0.995, 11)} \times \frac{0.94}{\sqrt{12}} = 0.55 \pm 3.11 \times 0.27 = 0.55 \pm 0.84 = (-0.29, 1.39)$$

برآورد تفاضل میانگین دو جامعه مستقل



- فرض کنید دو جامعه با میانگین‌های μ_1, μ_2 و واریانس‌های σ_1^2, σ_2^2 داریم، اگر تفاضل میانگین دو جامعه را با $\mu_1 - \mu_2$ نشان دهیم برای برآورد نقطه‌ای این پارامتر از برآوردگر $\bar{X}_1 - \bar{X}_2$ استفاده می‌کنیم.
- اما برای برآورد فاصله‌ای فرض کنید دو نمونه مستقل با حجم‌های n_1, n_2 از این دو جامعه انتخاب کردیم.
- فاصله اطمینان را در سه حالت ممکن بررسی می‌کنیم.
- (الف) واریانس‌ها معلوم
- (ب) واریانس‌ها نامعلوم ولی مساوی
- (ج) واریانس نامعلوم و نامساوی

فاصله اطمینان تفاضل میانگین دو جامعه (واریانس ها معلوم)

- در صورتی که واریانس دو جامعه معلوم باشد از فرمول زیر برای به دست آوردن فاصله اطمینان تفاضل میانگین دو جامعه استفاده می کنیم.

$$\bar{X}_1 - \bar{X}_2 - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- اگر فاصله عددی محاسبه شده از فواصل اطمینان فوق شامل صفر شود آنگاه با اطمینان $1 - \alpha$ می توان گفت که میانگین های دو جامعه مساوی هستند.

حالت اول: واریانس جامعه معلوم



متخصصین تغذیه ادعا می کنند در دو روش رژیم غذایی مختلف میزان کاهش وزن دارای توزیع نرمال با واریانس $\sigma_1^2 = 9, \sigma_2^2 = 16$ می باشد. اگر در یک نمونه $n_1 = 35$ تایی و $n_2 = 50$ به ترتیب متوسط وزن کاسته شده $\bar{X}_1 = 7.8, \bar{X}_2 = 9.2$ باشد، یک فاصله اطمینان برای تفاضل متوسط دو جامعه بدست آورید.

$$\begin{aligned} & \bar{X}_1 - \bar{X}_2 \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \\ & = 7.8 - 9.2 \pm 1.96 \times \sqrt{\frac{9}{35} + \frac{16}{50}} = -1.4 \pm 1.49 \\ & (-2.89, 0.09) \end{aligned}$$

حالت دوم: واریانس جامعه نا معلوم ولی مساوی



- اگر جدول زیر وزن نوزادان مادران سیگاری و وزن نوزادان با مادران غیر سیگاری نشان داده شده است، یک فاصله اطمینان برای تفاضل متوسط دو جامعه بدست آورید.

سیگاری	2.3	2.8	2.2	3.1	2.9	2.5	3.3	2.4	3.1
غیر سیگاری	3.3	2.9	3.6	2.8	3.4	3.1	2.6	3.9	2.7

$$\sum x_i = 24.6 \quad \sum x_i^2 = 68.5 \quad (n-1)S^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 68.5 - \frac{24.6^2}{9} = 1.26$$

$$\sum x_i = 28.3 \quad \sum x_i^2 = 90.53 \quad (n-1)S^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 90.53 - \frac{28.3^2}{9} = 1.54$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{1.26 + 1.54}{9 + 9 - 2} = 0.175 \quad S_p = 0.42$$

$$\bar{X}_1 - \bar{X}_2 \pm t_{\left(1-\frac{\alpha}{2}, n_1+n_2-2\right)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 2.73 - 3.14 \pm t_{(0.975, 16)} \times 0.42 \sqrt{\frac{1}{9} + \frac{1}{9}}$$

$$= -0.41 \pm 2.12 \times 0.42 \times 0.47 = -0.41 \pm 0.42 = (-0.01, 0.83)$$

حالت دوم: واریانس جامعه نا معلوم و نامساوی



- در این حالت از واریانس نمونه برای برآورد جامعه استفاده می شود ولی به دلیل مساوی نبودن واریانس ها مثل حالت قبل امکان یکی کردن واریانس ها و محاسبه واریانس آمیخته را نداریم. در این حالت فرمول محاسبه تفاضل میانگین ها در دو جامعه بر اساس فرمول زیر است:

$$\bar{X}_1 - \bar{X}_2 \pm t_{1-\frac{\alpha}{2}, df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- بطوری که مقدار درجه آزادی برای توزیع تی بر اساس واریانس ها در دو گروه و از فرمول زیر به دست می آید:

$$d.f. = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right]^2}{\frac{\left[\frac{S_1^2}{n_1} \right]^2}{n_1 - 1} + \frac{\left[\frac{S_2^2}{n_2} \right]^2}{n_2 - 1}}$$

مثال: در مطالعه برای بررسی اثر دو دارو بر روی مدت زمان بی حسی نتایج زیر به دست آمده است؟ یک فاصله اطمینان 90 درصد برای تفاضل میانگین دو نوع دارو با فرض نابرابری واریانس ها به دست آورید؟



65	35	40	55	60	45	A دارو
25	50	45	65	30	15	B دارو

$$\sum x_{i1} = 300 \quad \sum x_{i1}^2 = 15700 \quad S_1^2 = \frac{\sum x_{i1}^2 - \frac{(\sum x_{i1})^2}{n}}{n-1} = \frac{15700 - \frac{300^2}{6}}{5} = 140$$

$$\sum x_{i2} = 230 \quad \sum x_{i2}^2 = 10500 \quad S_2^2 = \frac{\sum x_{i2}^2 - \frac{(\sum x_{i2})^2}{n}}{n-1} = \frac{10500 - \frac{230^2}{6}}{5} = 336.67$$

$$d.f. = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right]^2}{\frac{\left[\frac{S_1^2}{n_1} \right]^2}{n_1-1} + \frac{\left[\frac{S_2^2}{n_2} \right]^2}{n_2-1}} = \frac{\left[\frac{140}{6} + \frac{336.67}{6} \right]^2}{\frac{\left[\frac{140}{6} \right]^2}{5} + \frac{\left[\frac{336.67}{6} \right]^2}{5}} = \frac{79.445^2}{108.89 + 629.70} = \frac{6311.51}{738.59} = 8.54 \cong 9$$

$$\bar{X}_1 - \bar{X}_2 \pm t_{1-\frac{\alpha}{2}, df} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 50 - 38.33 \pm 1.83 \times \sqrt{\frac{140}{6} + \frac{336.67}{6}} = 11.67 \pm 16.31 = (-4.64, 27.98)$$

آزمون فرض تساوی واریانس دو جامعه



- اگر $X_{11}, X_{12}, \dots, X_{1n_1}$ یک نمونه تصادفی از جامعه اول و $X_{21}, X_{22}, \dots, X_{2n_2}$ یک نمونه تصادفی از جامعه دوم باشند، آنگاه با استفاده از مشاهدات X_1 را به عنوان برآوردی از μ_1 ، $\bar{X}_1$ را به عنوان برآوردی از μ_2 ، S_1^2 را به عنوان برآوردی از σ_1^2 و S_2^2 را به عنوان برآوردی از σ_2^2 محاسبه می‌کنیم.

- حال تصمیم داریم دستگاه آزمون فرض زیر را با اطمینان $(1-\alpha)\%$ آزمون کنیم:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

- ملاک آزمون بصورت

$$F_c = \frac{S_1^2}{S_2^2} \approx F_{(1-\alpha; n_1-1, n_2-1)}$$

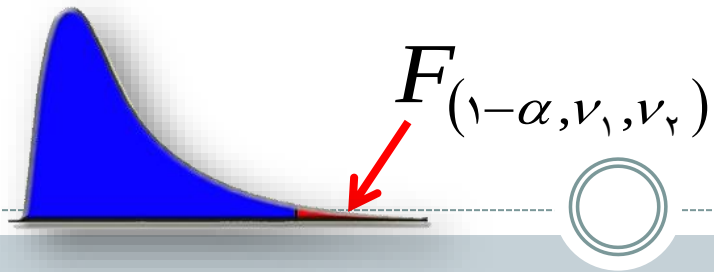
- برای این منظور از توزیع F و آماره مربوطه استفاده می‌کنیم.

• تحت فرض صفر انتظار می رود مقدار آماری F برابر ۱ باشد.

• برای مثال فرض کنید جراحی می خواهد بداند که آیا واریانس استحکام دو نوع نخ جراحی را یکی بداند؟ بدین منظور یک $n_1 = 16$ و $n_2 = 11$ از دو نوع نخ انتخاب کرده و با یک روش معین میزان استحکام نخ ها اندازه گیری می شود و مشاهده می شود که $S_1^2 = 4$ و $S_2^2 = 5$ می باشد.

• ملاک آزمون $F_c = \frac{S_1^2}{S_2^2} = \frac{5}{4} = 1.25$ به دست می آید که با در نظر گرفتن خطای ۰/۰۵

• با مقایسه با عدد بحرانی $F_{0.95}^{(10,15)} = 2.54$ فرض برابری واریانس در دو گروه پذیرفته می شود.



منحنی توزیع F

d2:
numerator

F Values for $\alpha = 0.05$

d_2	10	12	15	20	d_1 24	30
1	241.9	243.9	245.9	248.0	249.1	250.1
2	19.4	19.41	19.43	19.45	19.45	19.46
3	8.79	8.74	8.70	8.66	8.64	8.62
4	5.96	5.91	5.86	5.80	5.77	5.75
5	4.74	4.68	4.62	4.56	4.53	4.50
6	4.06	4.00	3.94	3.87	3.84	3.81
7	3.64	3.57	3.51	3.44	3.41	3.38
8	3.35	3.28	3.22	3.15	3.12	3.08
9	3.14	3.07	3.01	2.94	2.90	2.86
10	2.98	2.91	2.85	2.77	2.74	2.70
11	2.85	2.79	2.72	2.65	2.61	2.57
12	2.75	2.69	2.62	2.54	2.51	2.47
13	2.67	2.60	2.53	2.46	2.42	2.38
14	2.60	2.53	2.46	2.39	2.35	2.31
15	2.54	2.48	2.40	2.33	2.29	2.25
16	2.49	2.42	2.35	2.28	2.24	2.19
17	2.45	2.38	2.31	2.23	2.19	2.15
18	2.41	2.34	2.27	2.19	2.15	2.11

d1: denominator

فاصله اطمینان برای واریانس یک جامعه



- برای به دست آوردن فاصله اطمینان برای واریانس یک نمونه از فرمول زیر استفاده می کنیم.

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1, 1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1, \frac{\alpha}{2}}} \right)$$

- در این فرمول از توزیع کای اسکوئر استفاده می کنیم، این توزیع همانند توزیع تی استیودنت دارای پارامتری به اسم درجه آزادی است ولی بر عکس توزیع تی و نرمال نامتقارن بوده و فقط مقادیر مثبت را می پذیرد
- بهمین دلیل در توزیع کای اسکوئر صدک بالایی و پایینی قرینه نمی باشند و برای به دست آوردن مقادیر مربوطه باید از جدول مورد نظر استفاده نماییم.

مثال: داده های زیر میزان اضطراب کودکان قبل از ویزیت دندان پزشکی می باشد،
یک فاصله اطمینان 95 درصد برای واریانس به دست آورید؟



12	16	13	14	18	10	15	17	13
----	----	----	----	----	----	----	----	----

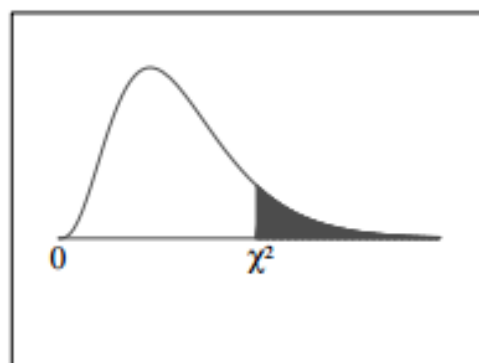
$$\sum x_i = 115$$

$$\sum x_i^2 = 1703$$

$$S^2 = \sum x_{i1}^2 - \frac{(\sum x_{i1})^2}{n} = 1703 - \frac{115^2}{9} = 233.56 \quad \chi_{0.025,9}^2 = 19.023 \quad \chi_{0.975,9}^2 = 2.7$$

$$\left(\frac{(n-1)S^2}{\chi_{n-1,1-\frac{\alpha}{2}}^2}, \frac{(n-1)S^2}{\chi_{n-1,\frac{\alpha}{2}}^2} \right) = \left(\frac{233.56}{2.18}, \frac{233.56}{17.54} \right) = (107.14, 13.32)$$

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_{\alpha}$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801

برآورد نسبت



- در مطالعاتی که متغیر تحت بررسی کیفی باشد در اینصورت متغیر حالات مختلف بخود می‌گیرد. مثلاً متغیرهایی که بصورت مثبت و منفی، زنده و مرده، سالم و بیمار، باسواد و بی‌سواد، شکست و پیروزی و باشد اینگونه متغیرها همانطور که قبلاً نیز ذکر گردیده است دو جمله‌ای می‌باشند و در اینجا هدف برآورد نسبت افرادی است که در جامعه ویژگی مورد نظر را دارند. به عنوان مثال وقتی که هدف برآورد نسبت نوزادان پسر در کل زایمانهایی که در طول یکسال اتفاق می‌افتد و یا هدف نسبت افرادی که در جامعه به یک بیماری مبتلا باشند یعنی در واقع برآورد شیوع بیماری در جامعه باشد. در اینصورت بر اساس یک نمونه n تایی، تعداد افرادی که در نمونه، ویژگی مورد نظر را داشته باشند شمارش گردیده (x) و نسبت واجدین شرایط در جامعه بر اساس نمونه انتخابی کسر $\left(\frac{x}{n}\right)$ می‌باشد. که یک برآورد نقطه‌ای برای P (نسبت جامعه) بر اساس نمونه به دست می‌آید.

برآورد فاصله‌ای نسبت (فاصله اطمینان)



- برای به دست آوردن يك فاصله اطمینان $(1-\alpha)\%$ برای نسبت جامعه از فرمول زیر استفاده می کنیم:

$$p \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}}$$

- که در این فرمول p نسبت افرادی که در نمونه ویژگی مورد نظر را دارند، می باشد.

مثال: در يك زایشگاه , 84 نوزاد در طول يك ماه متولد شده‌اند كه از این تعداد 37 دختر بوده‌اند. مطلوب است برآورد نقطه‌ای و فاصله ای با 95% برای نسبت دخترزایی در جامعه.



$$p = \frac{x}{n} = \frac{37}{84} = 0.44$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$$

$$Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96$$

$$p \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}}$$

$$0.44 \pm 1.96 \left(\frac{0.44(0.56)}{84} \right) =$$

$$0.44 \pm 0.106 =$$
$$(0.334, 0.546)$$

برآورد فاصله اطمینان تفاضل نسبت دو جامعه مستقل



- در مطالعه ای از بین 95 زن 25 نفر چاق و از بین 80 مرد 15 نفر چاق بودند می خواهیم یک فاصله اطمینان 95 درصدی برای اختلاف نسبت چاقی در زنان و مردان به دست آوریم.

$$(\bar{p}_1 - \bar{p}_2) \pm Z_{(1-\frac{\alpha}{2})} \times \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} =$$

$$\begin{aligned} \frac{25}{95} - \frac{15}{80} \pm 1.96 \sqrt{\frac{\frac{25}{95} \times \frac{70}{95}}{95} + \frac{\frac{15}{80} \times \frac{65}{80}}{80}} &= 0.26 - 0.19 \pm 1.96 \sqrt{0.002 + 0.002} \\ &= 0.07 \pm 1.96 \times 0.06 = 0.07 \pm 0.12 = (-0.05, 0.19) \end{aligned}$$

تمرین



- در یک مطالعه برای بررسی اثربخشی عصاره اسطوخودوس بر میزان اضطراب دانشجویان قبل از امتحان، قبل و بعد از استنشاق عصاره اندازه گیری شد. یک فاصله اطمینان 90 درصدی برای متوسط اختلاف اضطراب قبل و بعد از امتحان به دست آورید؟

قبل	17	16.5	18	18.2	19	16.7	15.5	14	15.9	16.3	18	17.5	15.5	16
بعد	16	15.5	19	18.5	17	16	16.5	12	15	14	17.5	16	16	14

تمرین



- در مطالعه ای بر روی اثربخشی میزان آموزش نحوه ی تغذیه صحیح در افراد دیابتیک نوع دوم بر میزان قند خون ناشتای آنان نتایج زیر به دست آمده است. یک فاصله اطمینان 95 درصدی برای متوسط اختلاف قند خون ناشتا در دو گروه به دست آورید؟

کنترل	110	115	120	113	117	125	108	105
درمان	102	110	113	107	100	120	98	95

تمرین



- در مطالعه ای از بین 200 دانشجوی دختر 60 نفر متاهل و از بین 50 دانشجوی پسر 10 نفر متاهل بودند. یک فاصله اطمینان 90 درصدی برای اختلاف نسبت تاهل در دختران و پسران به دست آورید؟

$$(\bar{p}_1 - \bar{p}_2) \pm Z_{(1-\frac{\alpha}{2})} \times \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}} =$$

$$\frac{60}{200} - \frac{10}{50} \pm 1.64 \times \sqrt{\frac{\frac{60}{200} \times \frac{140}{200}}{200} + \frac{\frac{10}{50} \times \frac{40}{50}}{50}} =$$

$$0.1 \pm 1.64 \times 0.065 = 0.1 \pm 0.107 = (0.217, 0.07)$$

محاسبه شانس Odds



- شانس یک روش برای بیان احتمال است. شانس وقوع یک پیشامد برابر است با احتمال وقوع آن به احتمال عدم وقوع آن .

$$odds = \frac{p}{1-p} \text{ or } p = \frac{Odds}{1+odds}$$

- برای مثال اگر احتمال داشتن شغل برای یک مرد 60 درصد باشد، بدین معنی است که شانس داشتن شغل برای او 1.5 است.

نسبت شانس (Odds Ratio)



$$OR = \frac{p_1 / 1 - p_1}{p_2 / 1 - p_2}$$

- نسبت دو شانس بر روی هم را نسبت شانس گویند.

- برای مثال اگر شانس داشتن شغل برای مردان 1.5 و برای زنان 0.5 باشد. نسبت شانس داشتن شغل برای مردان نسبت به زنان 3 به دست می آید و برعکس شانس داشتن شغل برای زنان نسبت به مردان 1/3 خواهد بود.
- **برای تفسیر نسبت شانس برای داشتن شغل مردان به زنان**
- اگر حاصل عددی بین صفر و یک شد مثلاً 1.4 می گوئیم مردان 40 درصد شانس بیشتری نسبت به زنان برای داشتن شغل دارند.
- اگر حاصل عددی بزرگتر از دو شد مثلاً 2.7 می گوئیم مردان 2.7 برابر شانس بیشتری نسبت به زنان برای داشتن شغل دارند.
- اگر حاصل کمتر از یک شد مثلاً 0.7 می گوئیم مردان 30 درصد شانس کمتری نسبت به زنان برای داشتن شغل دارند.

Properties of the Odds Ratio



- نسبت شانس می تواند هر عدد غیر منفی باشد.
- زمانی که دو متغیر X و Y از هم مستقل هستند، نسبت شانس یک است.
- نکته: X و Y دو حالتی هستند.

جمع	خیر	بله	سیگار سکته
25	5	20	بله
25	15	10	خیر
50	20	30	جمع

- در جدول زیر می توان نسبت شانس سکته در افراد سیگاری به غیر سیگاری را بصورت زیر به دست آورد:

- احتمال سکته کردن در افراد سیگاری: $20/30$

- احتمال سکته کردن در افراد غیر سیگاری $5/20$

- نسبت شانس سکته افراد به سیگاری به غیر سیگاری

$$OR = \frac{\frac{20}{30}}{\frac{5}{20}} = \frac{20/10}{5/15} = \frac{20 \times 15}{5 \times 10}$$

محاسبه فاصله اطمینان برای نسبت شانس



- برای محاسبه فاصله اطمینان برای نسبت شانس در یک جدول 2 در 2 مانند جدول زیر از فرمول زیر استفاده می کنیم:

جنسیت فشارخون	مرد	زن
دارد	a	c
ندارد	b	d

$$\ln(OR) \pm Z_{1-\alpha/2} \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

- فاصله اطمینان برای نسبت شانس داشتن فشار خون مردها به زنها

$$OR = \frac{ad}{bc}$$

با توجه به اطلاعات جدول داده‌شده یک فاصله اطمینان 95 درصدی برای نسبت شانس داشتن فشار خون مردها به زنها به دست آورید؟

زن	مرد	جنسیت فشارخون
20	15	دارد
80	45	ندارد

$$OR = \frac{ad}{bc} = \frac{15 \times 80}{20 \times 45} = \frac{4}{3} = 1.33$$

$$\ln(OR) \pm Z_{1-\alpha/2} \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} = 0.29 \pm 1.96 \sqrt{\frac{1}{15} + \frac{1}{45} + \frac{1}{20} + \frac{1}{80}} =$$

$$0.29 \pm 1.96 \times 0.39 = 0.29 \pm 0.76 = (1.05, -0.47)$$

- این فاصله اطمینان برای لگاریتم پایه طبیعی است نسبت شانس است برای آنکه فاصله اطمینان برای نسبت شانس به دست آوریم باید توان نمایی را برای مقادیر به دست آمده محاسبه نماییم.
- هنگامی که این فاصله یک را در بر می‌گیرد بدان معنی است که شانس داشتن فشار خون در دو گروه با هم اختلاف معنی داری ندارد.

$$(e^{1.05}, e^{-0.47}) = (2.86, 0.625)$$

فاصله اطمینان برای ضریب همبستگی پیرسون



- همانطور که قبلاً گفته شد برای ضریب همبستگی پیرسون از فرمول زیر استفاده می کنیم:

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n} \right) \left(\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right)}}$$

- ونشان دادیم که وقتی تبدیل $w = \frac{1}{2} \ln \frac{1+r}{1-r}$ را بکار بردیم این آماره دارای توزیع نرمال با انحراف معیار $\frac{1}{\sqrt{n-3}}$ می باشد پس برای به دست آوردن فاصله اطمینان برای آن از رابطه زیر استفاده می کنیم:

$$w \pm Z_{1-\alpha/2} \left(\frac{1}{\sqrt{n-3}} \right)$$

در مطالعه ای برای بررسی رابطه بین بهره هوشی و پیشرفت تحصیلی نتایج زیر به دست آمده است یک فاصله اطمینان 95 درصد برای ضریب همبستگی پیرسون به دست آورید؟



127	116	123	118	125	120	X= بهره هوشی
17.3	15.8	16.2	15.4	16.7	14.8	Y= معدل

$$\sum x_i = 729$$

$$\sum x_i^2 = 88663$$

$$\sum x_i y_i = 11703.2$$

$$\sum y_i = 96.2$$

$$\sum y_i^2 = 1546.46$$

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right) \left(\sum y_i^2 - \frac{(\sum y_i)^2}{n} \right)}} = \frac{11703.2 - \frac{729 \times 96.2}{6}}{\sqrt{\left(88663 - \frac{(729)^2}{6} \right) \left(1546.46 - \frac{(96.2)^2}{6} \right)}} = \frac{14.9}{\sqrt{89.5 \times 4.05}} = 0.78$$



$$w = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) = 0.5 \times \ln(8.1) = 0.5 \times 2.1 = 1.045$$

$$w \pm Z_{1-\alpha/2} \left(\frac{1}{\sqrt{n-3}} \right) = 1.045 \pm 1.96 \times \frac{1}{\sqrt{3}} = 1.045 \pm 1.13 = (-0.085, 2.058)$$

- حال باید تبدیل زیر را بکار ببریم تا فاصله اطمینان برای ضریب همبستگی پیرسون به دست آید.

$$r = \frac{\exp(2w) - 1}{\exp(2w) + 1}$$

- که در این مثال بصورت زیر به دست می آید: (0.99 و -0.035)
- چون فاصله اطمینان صفر را در بر گرفته می گوئیم با اطمینان 95 درصد رابطه ای بین بهره هوشی و معدل وجود ندارد.