بسمه تعالی

آمار و SPSS

راضیه عبدالرحمن زده

raziehabdolrahmanzadeh@gmail.com

زمستان ۱۴۰۱

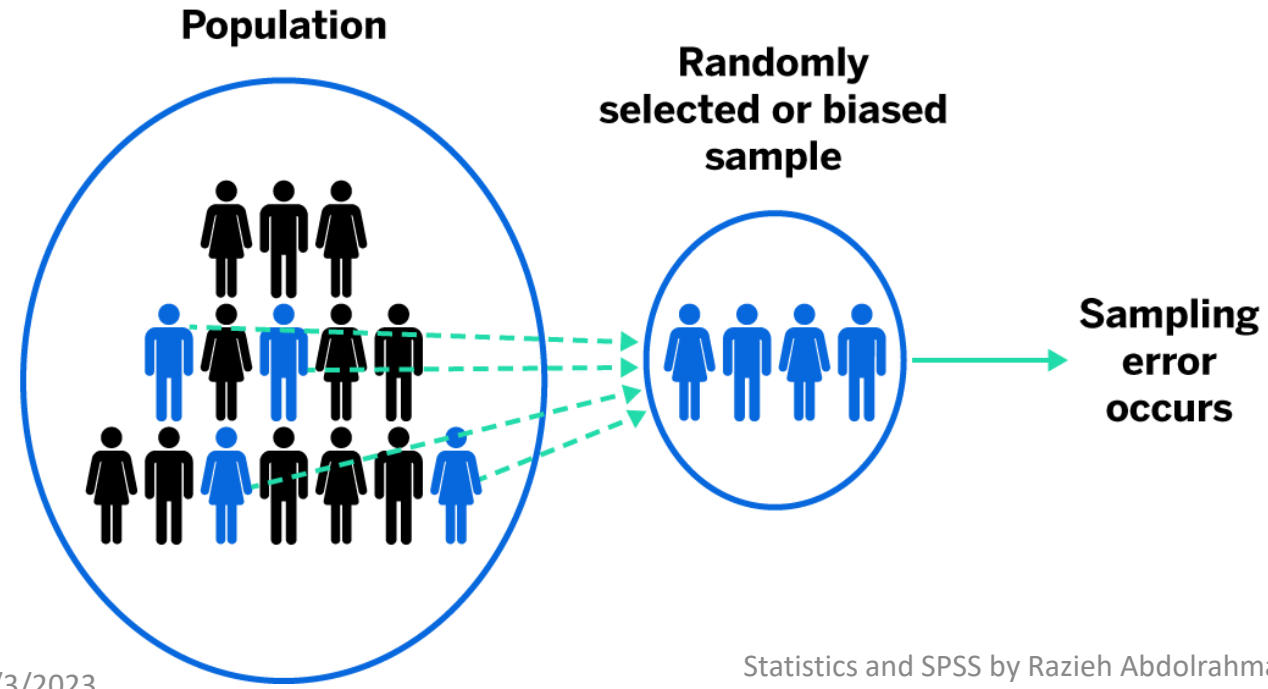
آمار استنباطی:

- در مورد وجود یا عدم وجود ارتباط بین دو پدیده، تشابه یا عدم تشابه دو پدیده و یا موارد مشابه در مطالعات تصمیم گیری شود پای آمار تحلیلی به وسط می آید.
- قدرت استباط یا برآورد تحت تأثیر روش نمونه گیری قرار دارد.
- آمار استنباطی با دو دسته از مسائل سروکار دارد: ابتدا به برآورد و سپس به آزمون می پردازد.

Statistics and SPSS by Razieh Abdolrahmanzadeh
raziehabdolrahmanzdeh@gmail.com

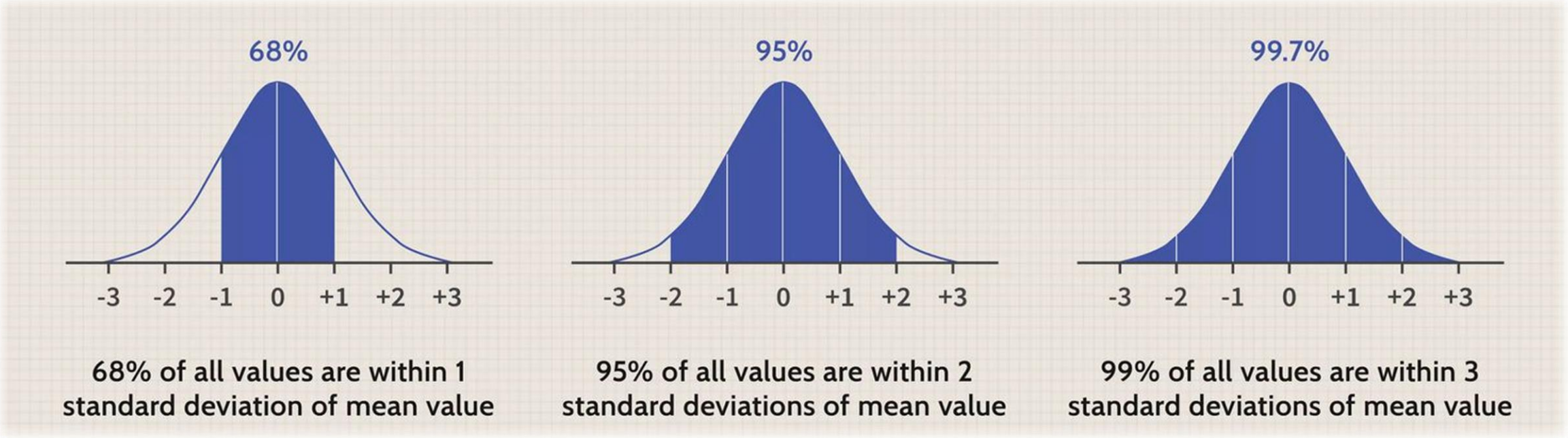
خطا نمونه گیری:

- چنانچه از جامعه نمونه های متفاوتی را به صورت تصادفی انتخاب کنیم ملاحظه خواهیم کرد که همه آن ها دارای ویژگی های یکسانی نیستند و حتی گاهی مشابه جامعه هم نیستند که به علت وجود خطا نمونه گیری است.
- خطا نمونه گیری ضرورتاً نتیجه اشتباه نمونه گیری نیست و این خطا ممکن است به علت اختلاف های فاحش در بین آزمودنی ها باشد.



قضیه حد مرکزی (central limit theorem):

اگر از یک جامعه به صورت تصادفی نمونه های زیادی را با اندازه های مساوی انتخاب کنیم و میانگین این نمونه ها را محاسبه کنیم، توزیع این میانگین ها یک توزیع طبیعی خواهد شد و میانگین میانگین های نمونه های انتخاب شده تقریباً برابر با میانگین جامعه خواهد بود.



- میانگین یک نمونه با ۶۸,۲۶٪ اطمینان بین ۱± انحراف استاندارد از میانگین جامعه قرار خواهد گرفت.
- میانگین یک نمونه با ۹۵٪ اطمینان بین ۱,۹۶± انحراف استاندارد از میانگین جامعه قرار خواهد گرفت.
- میانگین یک نمونه با ۹۹٪ اطمینان بین ۲,۵۸± انحراف استاندارد از میانگین جامعه قرار خواهد گرفت.
- میانگین یک نمونه با ۹۹,۷۳٪ اطمینان بین ۳± انحراف استاندارد از میانگین جامعه قرار خواهد گرفت.

خطای استاندارد میانگین (standard error of mean):

انحراف استاندارد توزیع نظری میانگین ها شاخصی است که به وسیله خطا نمونه گیری اندازه گیری می شود و آن را خطا استاندارد میانگین می نامند.

$$SE = \frac{\sigma}{\sqrt{n}}$$

σ ← Standard deviation

n ← Number of samples

هرچه انحراف معیار قد افراد در جمعیت اصلی بیشتر باشد، احتمالاً انحراف معیار میانگین هایی که در نمونه گیری از آن حاصل می شود بیشتر خواهد بود.

اگر اندازه نمونه ما کوچکتر باشد احتمال اینکه وجود اعضا پرت در آن باعث پرت به دست آمدن میانگین آن نمونه می شود بیشتر از حالتی است که اندازه نمونه زیاد باشد.

آزمون فرضیه:

- یک فرآیند استنتاجی است که هدف آن برآورد پارامترهای جامعه ای است که نمونه مورد مطالعه از آن استخراج شده است.
- آزمون فرضیه با بیان فرضیه های آماری آغاز می شود.
- فرضیه آماری جمله ای است که پیرامون ویژگی های جامعه بیان می شود و امکان دارد درست نباشد ولی پژوهشگر صرفاً به خاطر برقرار کردن یک شرایط قابل آزمایش آن را بیان می کند.
- دو نوع فرضیه آماری داریم:

۱. فرض صفر (H_0)

۲. فرض جانشین یا مقابل (H_1)

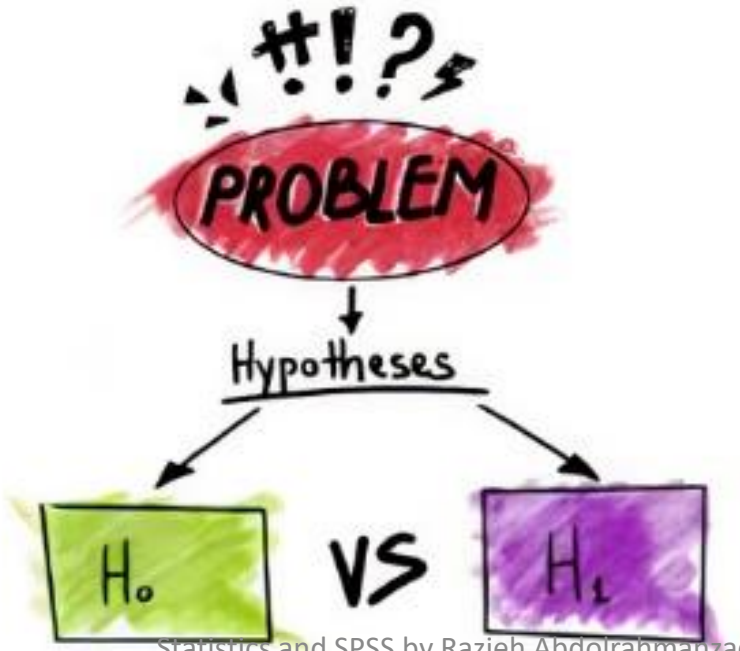
فرض صفر (Null Hypothesis):

بین پارامتر های مورد مطالعه اختلاف یا ارتباط معناداری وجود ندارد.

فرض جانشین یا مقابل(Alternative Hypothesis):

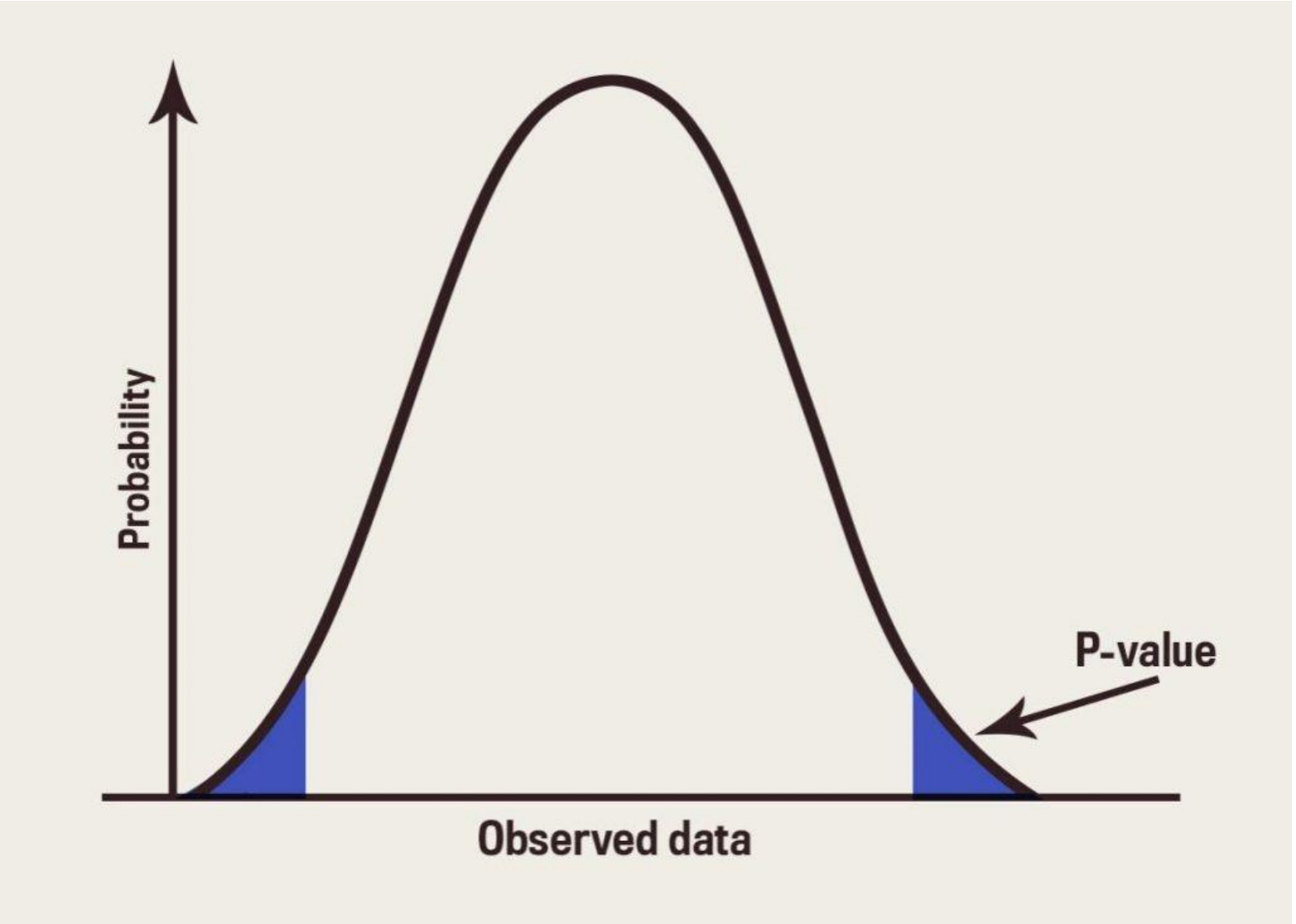
مخالف فرض صفر است. بین پارامترهای مورد مطالعه اختلاف یا ارتباط معناداری وجود دارد.

- اکثر مواقع با فرضیه پژوهشی مطابقت دارد و انتظار پژوهشگر را پیرامون نتایج آتی پژوهش بیان می کند.



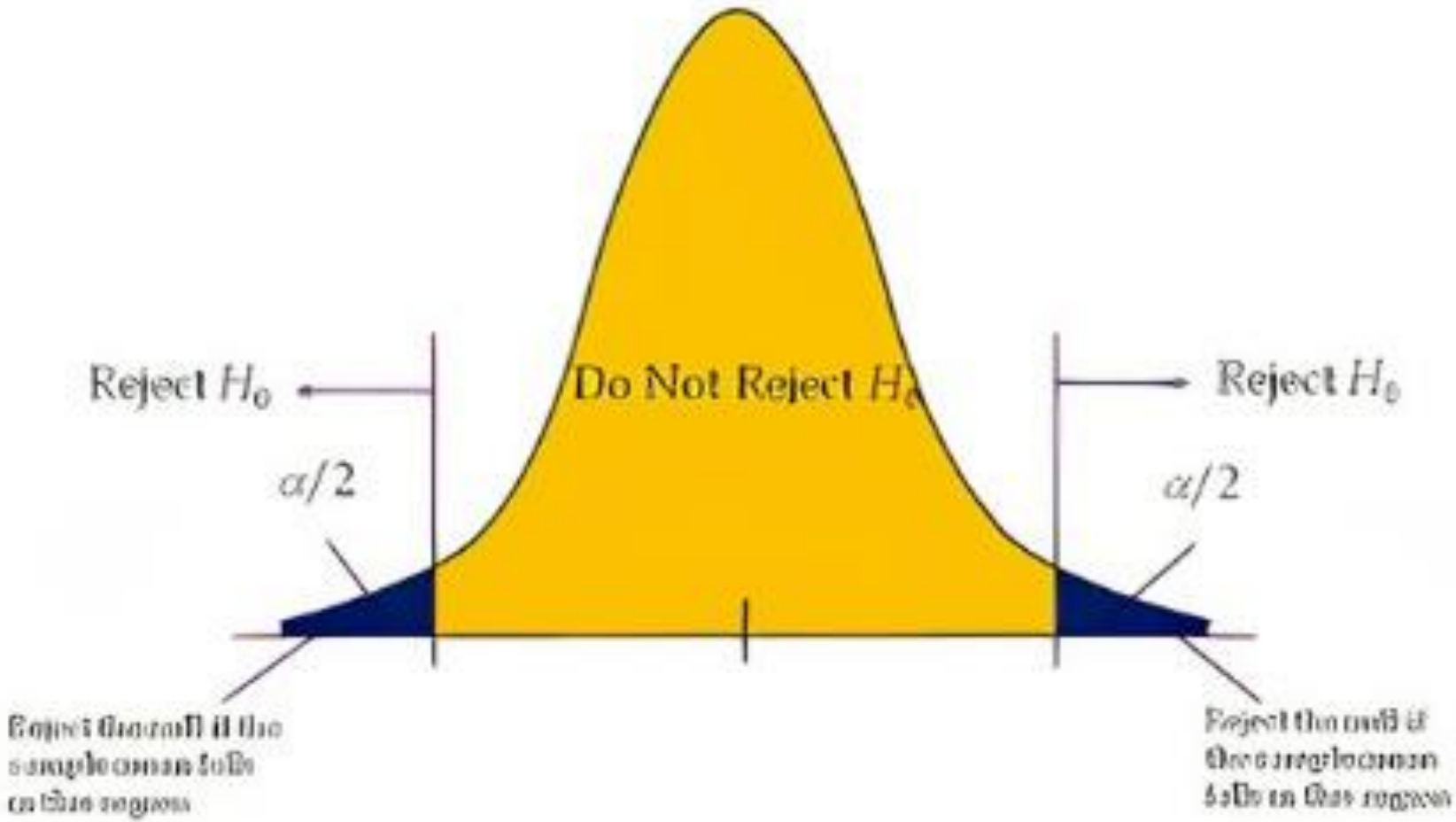
p-value:

- با کمک محاسبه p-value توانستیم راهکاری را برای رد کردن یا رد نکردن H_0 داشته باشیم.
- احتمال رخ دادن اتفاق هایی که حداقل به میزان تفاوت مشاهده شده با حد مورد انتظار اختلاف داشته باشند.
- p-value برای اوج ایده آل برابر یک است اما هر چه اتفاقی که رخ می دهد از ایده آل ما دورتر باشد p-value به صفر نزدیک تر می شود.
- طبق سنت تاریخی در آمار به طور پیش فرض عدد ۰,۰۵ برای مرز p-value در معنی دار بودن یا نبودن اختلاف در نظر گرفته می شود.


Probability
P-value
Observed data

Significance:

- گاهی اوقات به جای استفاده از کلمه p-value از واژه significance استفاده می شود که به معنادار بودن تبعیت از H_0 اشاره دارد که دقیقاً برابر p-value است.
- اگر p-value در تستی زیادتر از حد مرزی(نزدیک به یک) باشد می گوییم به طور معناداری از توزیع H_0 تبعیت می کند پس H_0 را نمی توان رد کرد.
- اگر p-value در تستی کمتر از حد مرزی (نزدیک به صفر) باشد، می گوییم تبعیت آن از H_0 معنی دار نیست پس H_0 رد می شود.


Reject H_0
Do Not Reject H_0
Reject H_0
$\alpha/2$
$\alpha/2$

خطا نوع اول (Type one error):

- ممکن است رخداد ما واقعاً از توزیع موردنظر ما تبعیت کرده باشد ولی ما به اشتباه آن را رد کنیم.
- خطای نوع اول زمانی اتفاق می افتد که فرض صفر درست باشد ولی ما آن را رد کرده باشیم.
- خطای نوع اول را با آلفا نشان می دهند. در نتیجه هنگامی که در یک آزمون آماری مقدار آلفا تعیین می شود، در حقیقت احتمال ارتکاب خطا نوع اول تعیین می گردد.
- این خطا دقیقا برابر است با همان ۵٪ مساحتی که زیر نمودار به دلیل دورتر بودن از مرز مورد نظرمان H_0 را برایشان رد می کنیم.

خطا نوع دوم یا بتا (Type two error):

- واقعه ای را که در حقیقت عضو توزیع ما نیست را صرفاً به دلیل آن که از نظر محاسبه p-value در محدوده عدم رد H_0 قرار دارد، به اشتباه عضو توزیع مان در نظر بگیریم و H_0 را اشتباهاً برای آن رد نکنیم.

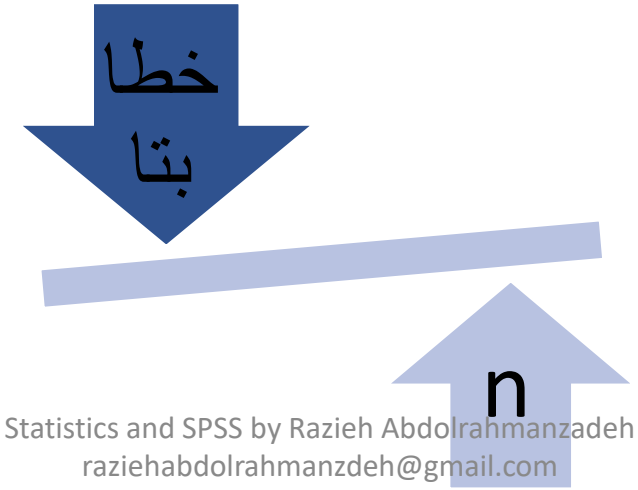
احتمال ارتکاب به خطا نوع دوم به چهار عامل زیر بستگی دارد:

۱. سطح معنادار بودن: تغییر در مقدار آلفا در تعیین محدوده ناحیه رد که برای آزمون آماری انتخاب می شود، تأثیر دارد.

۲. اندازه تأثیر متغیر مستقل: بین تأثیر و شدت متغیر مستقل و خطا نوع دوم رابطه معکوس وجود دارد.

۳. مقدار پراکندگی موجود در متغیر وابسته: برای کاهش احتمال ارتکاب خطا دوم و در نتیجه افزایش توان آزمون پژوهشگر باید سعی کند پراکندگی متغیر وابسته را به حداکثر مقدار برساند.

۴. اندازه یا حجم نمونه: با افزایش نمونه خطای آلفا و بتا همزمان کاهش پیدا خواهد کرد.

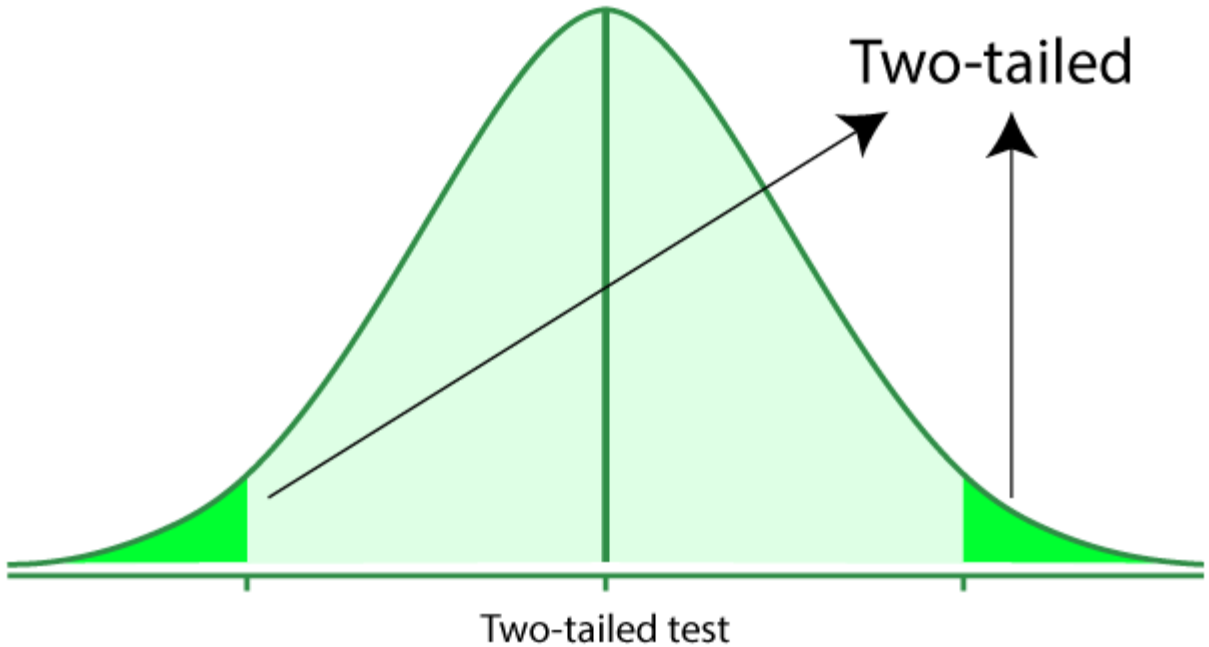


		تصمیم در مورد فرض صفر مطابق آزمون آماری	
		عدم رد فرض صفر	رد فرض صفر
واقعیت در مورد پارامتر جمعیت مطابق با فرض صفر	درست	تصمیم درست	خطای نوع اول
	غلط	خطای نوع دوم	تصمیم درست

آزمون دو دامنه (two tailed test):

زمانی آزمون بدون جهت است که جهت تأثیر متغیر مستقل بر متغیر وابسته معلوم نباشد. به این گونه آزمون ها دو دامنه یا دو طرفه گویند؛ زیرا در این نوع آزمون ها دو دامنه برای رد فرضیه صفر وجود دارد.

وقتی آزمون دو دامنه به کار برده می شود الفا به دو قسمت مساوی تقسیم می شود.

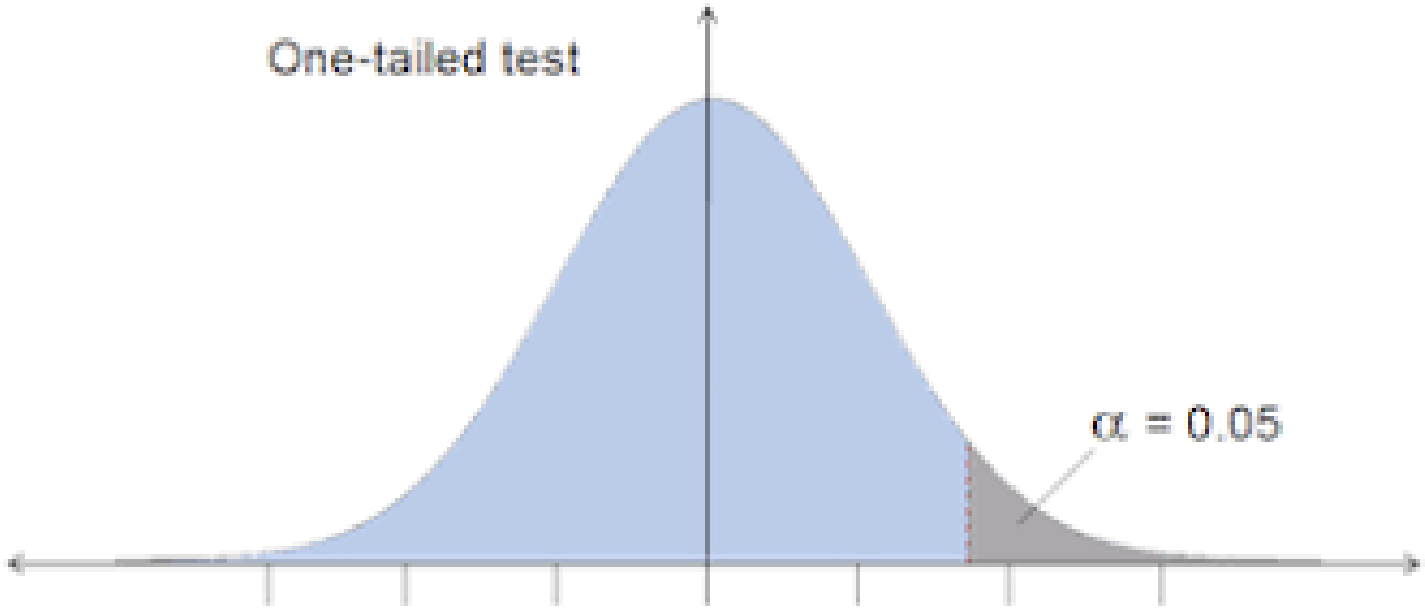


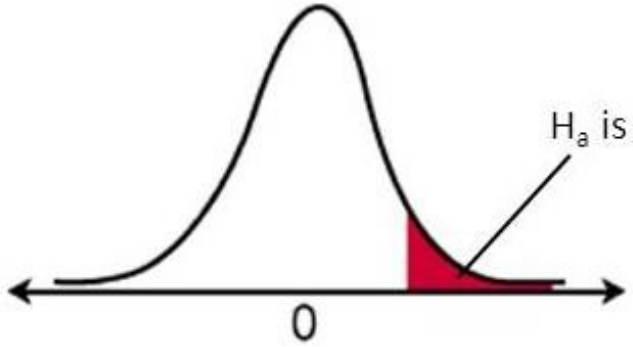
آزمون یک دامنه (one tailed test):

در برخی شرایط تحقیق ممکن است قبل از اجرای تحقیق، درباره جهت اختلاف بین شاخص های آماری مورد مطالعه تصمیم گیری کند.

فرضیه صفر تغییر نمی کند و تنها فرضیه خلاف است که دستخوش تغییر می شود.

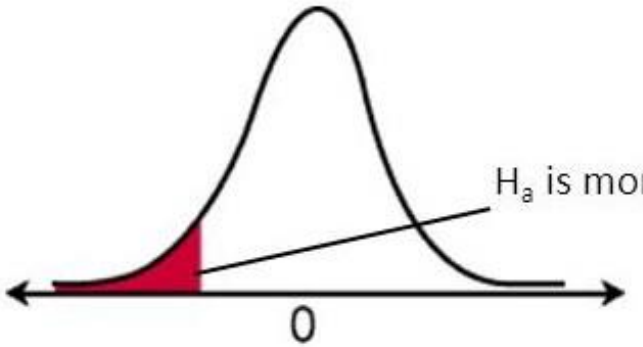
در آزمون های یک دامنه فقط یک ناحیه یا دامنه جهت رد فرض صفر وجود دارد.





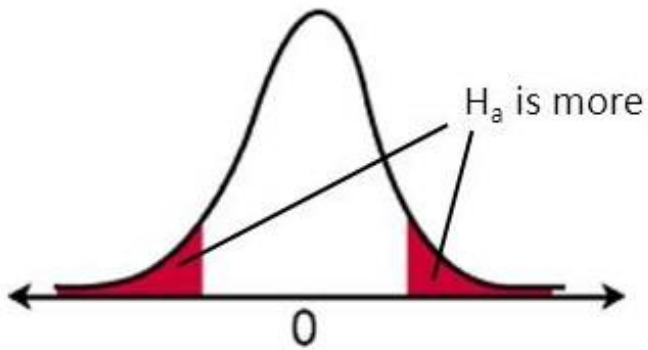
Right-tail test

$H_a: \mu > \text{value}$



Left-tail test

$H_a: \mu < \text{value}$



Two-tail test

$H_a: \mu \neq \text{value}$

❖ از این لحظه به بعد قبل از انجام هر تست آماری چهار موضوع زیر را برای خود مشخص کنید:

۱. فرضیه های H_0 و H_1 چیست؟

۲. تست ما One tailed هست یا two tailed؟

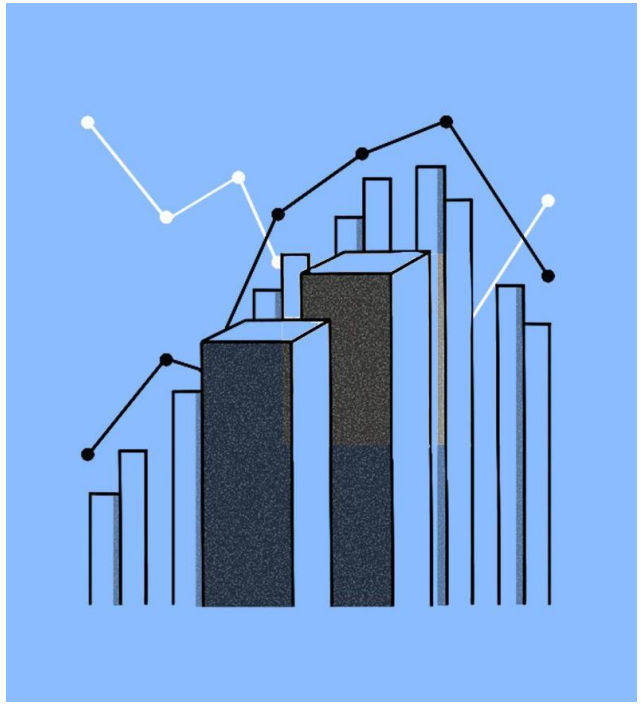
۳. چند درصد خطا نوع اول برای قضاوت در نظر می گیریم؟

۴. از چه تستی باید استفاده کنیم؟

:z-test

یک آزمون آماری برای تعیین اینکه میانگین دو جمعیت تفاوت معناداری باهم دارند یا خیر زمانی که واریانس معلوم و حجم نمونه بزرگ است.

از آن جا که در چنین تستی از توزیع و آماره استاندارد z استفاده می شود به آن z test گفته می شود.



:OneSample z-test

مقایسه میانگین یک نمونه با میانگین یک جمعیت در صورت مشخص بودن انحراف استاندارد جمعیت

$$Z\ Test = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

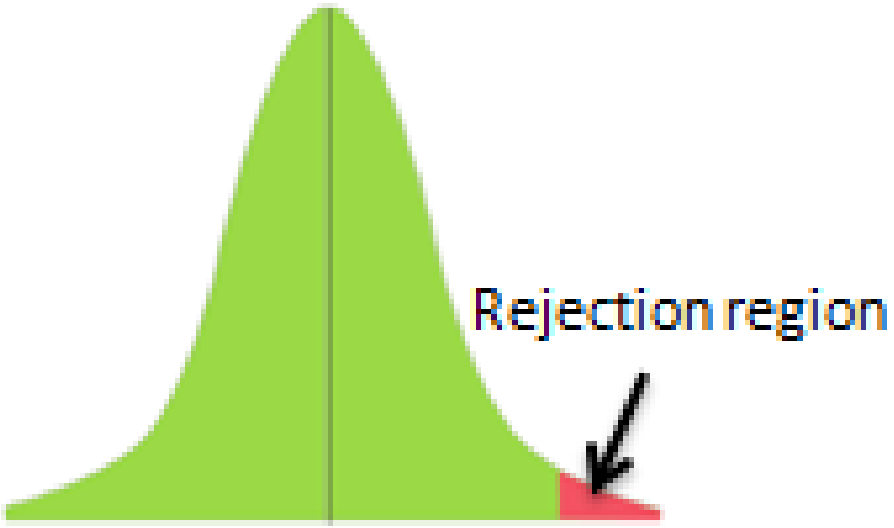
مثال:

نمرات کل جمعیت شرکت کننده در آزمون بین المللی CRE انحراف معیار ۸,۵ دارد.

دانشگاه روانشناسی UW دوست دارد نمره آزمون GRE دانشجویان پذیرفته شده در آن دانشگاه، بیش از ۲۱۰ باشد. میانگین نمرات آزمون GRE برای ۴۲ دانشجوی پذیرفته شده امسال ۲۱۲,۷۹ بود.

الف- بر این اساس تعیین کنید که آیا نمره این دانشجویان به طور معناداری بیش از ۲۱۰ بوده است یا نه؟

ب- حداقل میانگین نمره ۴۲ دانشجو که با خطا ۵٪ به طور معناداری بیشتر از ۲۱۰ تلقی می شود، چند است؟



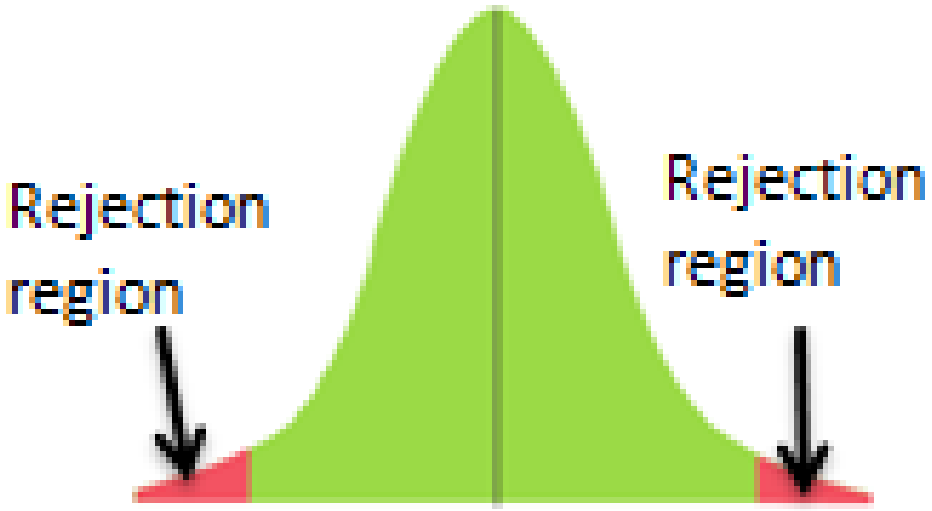
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

مثال:

فرض کنید شما یک شرکت دارویی تأسیس کرده اید و در آن دارویی را تولید کرده اید که قرار است IQ را افزایش دهد. می دانید که میانگین و انحراف معیار IQ در جمعیت عامه مردم به ترتیب ۱۰۰ و ۱۵ است. شما دارویتان را روی ۳۶ فرد تصادفی تست می کنید و میانگین IQ برابر با ۱۰۲,۹۶ را از آن ها به دست می آورید. با خطا ۱٪:

الف- مشخص کنید که آیا IQی افراد که دارو گرفته اند به طور معناداری با متوسط IQ عامه مردم تفاوت دارند یا نه؟

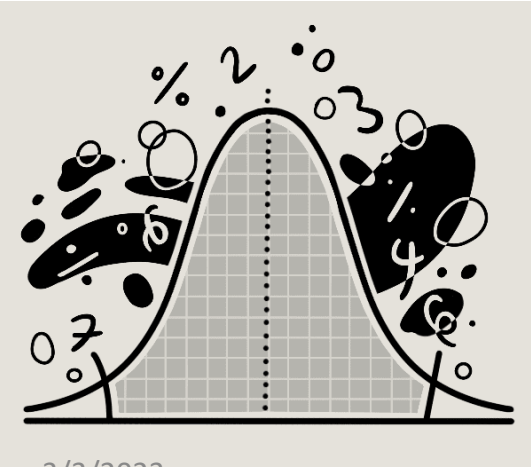
ب- میانگین نمونه ۳۶ تایی شما باید خارج چه بازه ای باشد تا بگوییم IQی آن ها با عامه مردم تفاوت دارد؟



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

پیشبینی میانگین جمعیت و محدوده اطمینان (Confidence Interval):

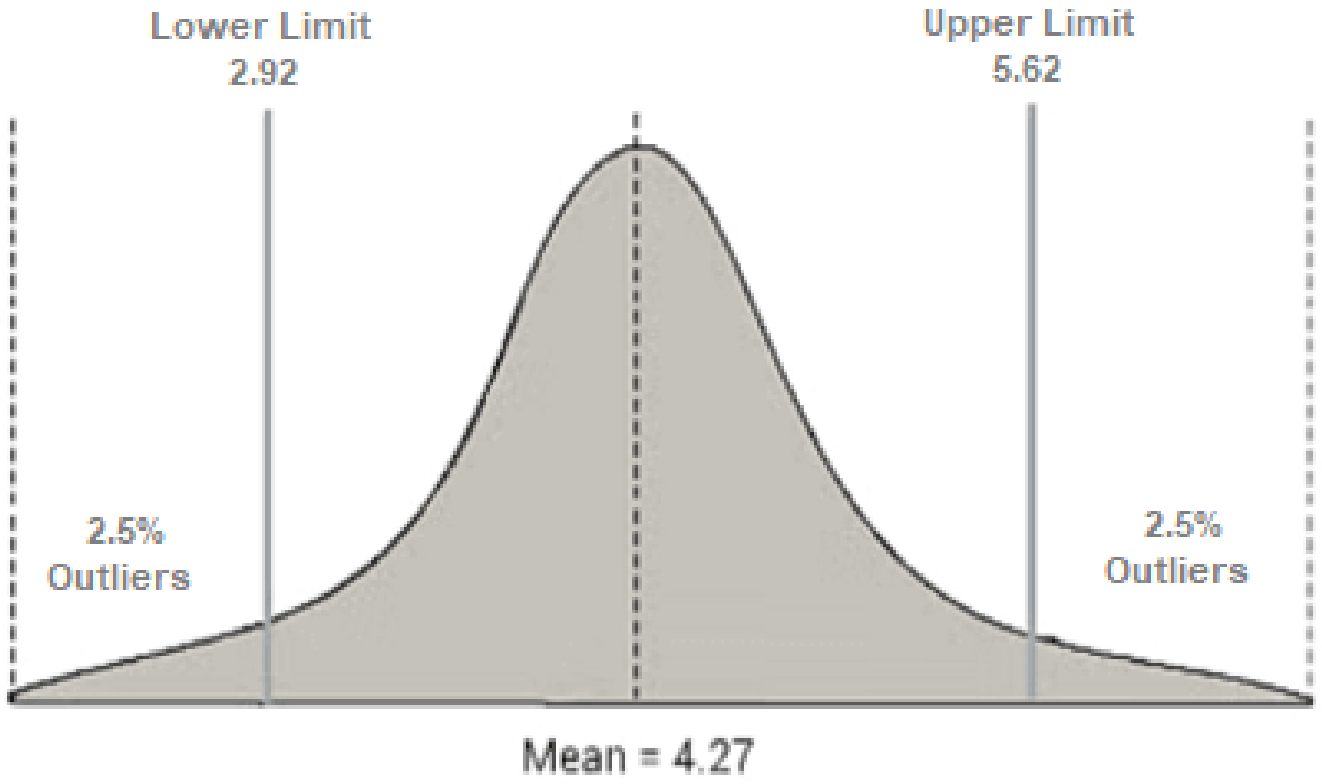
- گاهی اوقات وقتی از یک جمعیت نمونه گیری می کنیم، میانگین جمعیت اصلی را نداریم و میخواهیم آن را بر اساس نمونه ای که گرفتیم پیشبینی کنیم.
- بهترین کاری که برای پیشبینی میانگین از دستمان برمی آید آن است که بر اساس میانگین نمونه خودمان محدوده ای را مشخص کنیم که به احتمال زیاد میانگین جمعیت اصلی در آن قرار خواهد داشت.
- چنین محدوده ای اصطلاحاً confidence interval یا محدوده اطمینان برای میانگین جمعیت نامیده می شود.



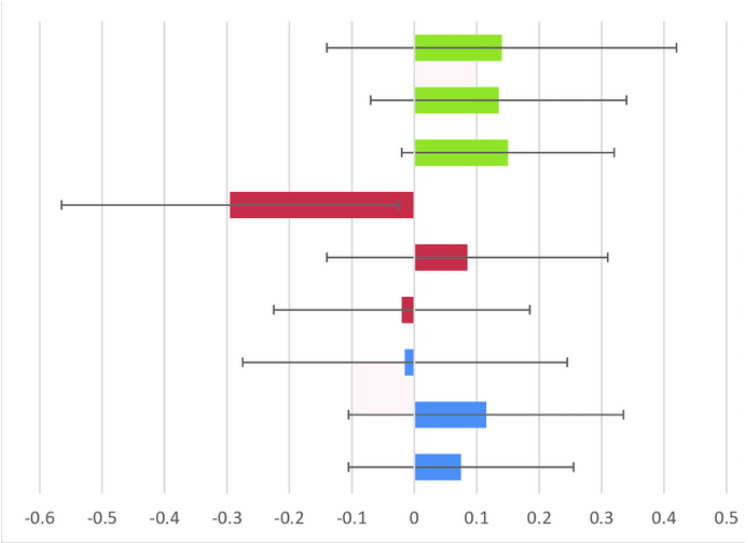
$$\text{Confidence Interval} = \bar{X} \pm Z \times \frac{\sigma}{\sqrt{n}}$$

Confidence Interval

95% Chance your population mean
will fall between 2.92 and 5.62



- اگر چندین و چندبار از یک جمعیت نمونه گیری کنیم هر بار میانگین جدیدی به دست می آید و بر اساس هر میانگین جدید CI جدیدی به دست خواهیم آورد که این CIها با هم همپوشانی دارند.
- اگر خطا مدنظر در محاسبه CIها ۵٪ باشد، میانگین واقعی جمعیت در بازه مشخص شده توسط ۹۵٪ این CIها قرار می گیرد.



- در z-testها می توان به جای محاسبه p-value از CI برای مقایسه میانگین نمونه با میانگین جمعیت استفاده کرد.

t-test:

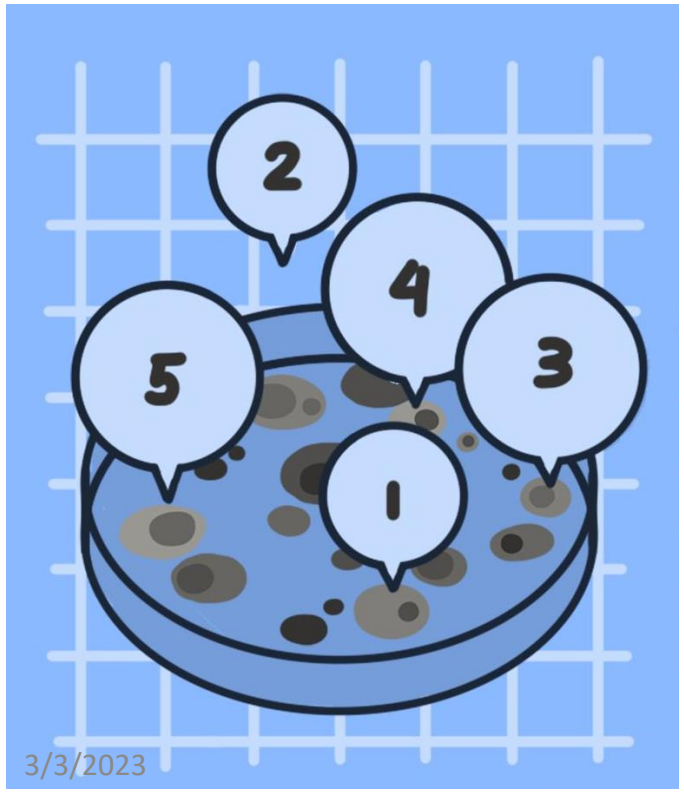
در z-test انحراف معیار جمعیت اصلی را داشتیم ولی این تقریباً غیر ممکن است و ما فقط می توانیم انحراف معیار همان اعضا نمونه ای را که گرفتیم(s) را حساب کنیم که لزوماً با انحراف معیار جمعیت یکسان نیست.

$$t = \frac{(x - \mu)}{\frac{s}{\sqrt{n}}}$$

قسمت اول t-test کاملا شبیه z-test است ولی به جای اینکه score را در توزیع استاندارد z بررسی کنیم، باید در توزیع t بررسی کنیم.

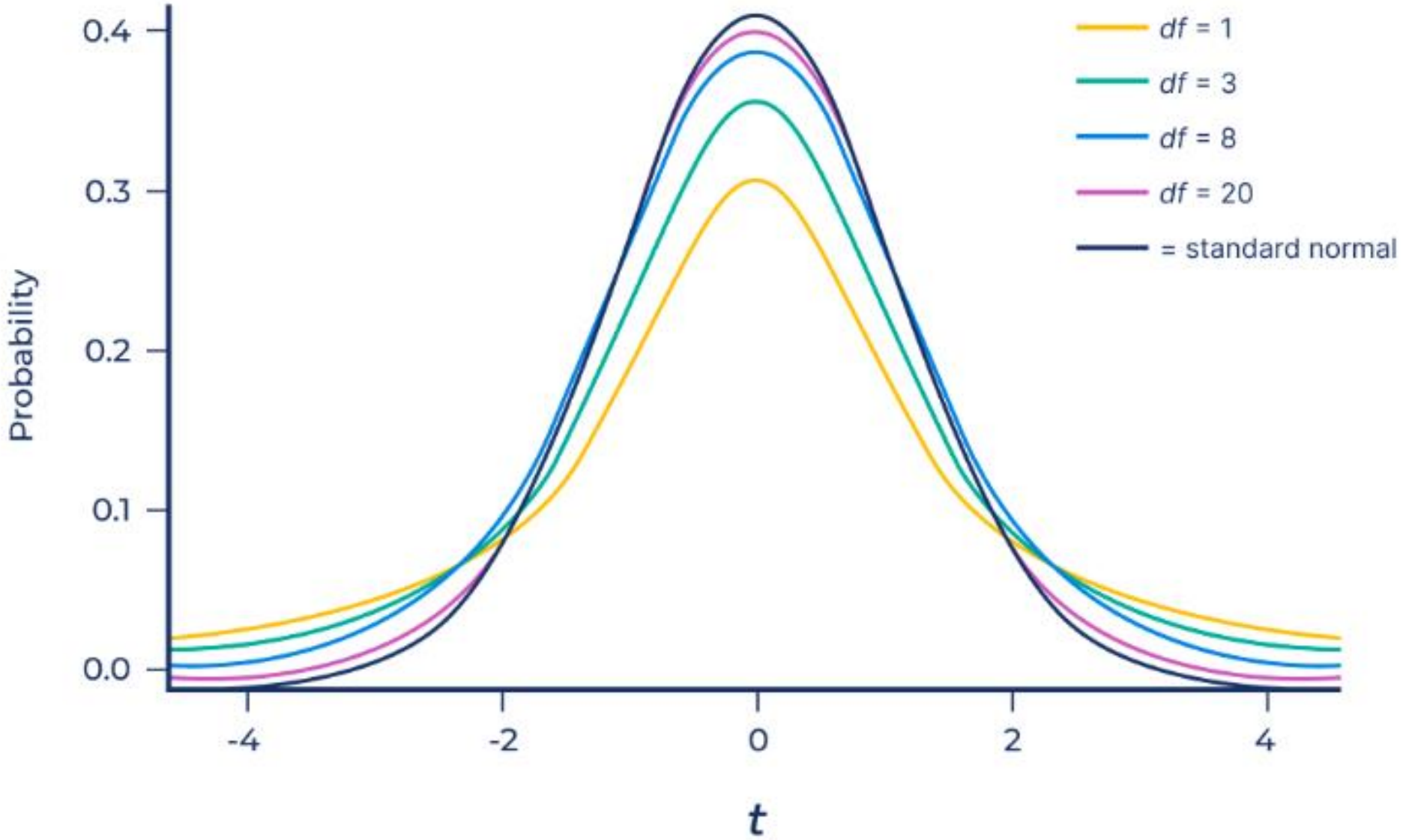
درجه آزادی (Degree of Freedom):

- تعداد متغیرهایی در یک مدل که می توانند به راحتی تغییر کنند.
- درجه آزادی است که تعداد حالات ممکن و مشخصات آماری نمونه ها را معین می کند.
- در اکثر کارهای آماری که با نمونه گیری انجام می شود باید به جای تعداد اعضا نمونه(n) از درجه آزادی آن نمونه استفاده کرد.
- در اکثر شرایط درجه آزادی برابر n-1 است.



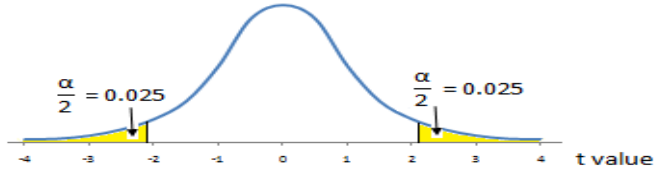
توزیع t:

- در واقع مثل توزیع استاندارد نرمال است با میانگین صفر و انحراف استاندارد یک.
- برخلاف توزیع استاندارد که در تمام شرایط همواره یک مختصات ثابت داشت، توزیع t برحسب اینکه درجه آزادی نمونه مورد بررسی ما چند است، مختصات و شکل منحصر به فرد دارد.
- شکل نمودار وقتی درجه آزادی کم است(اندازه نمونه کوچک است) تفاوت زیادی باهم دارند.
- هرچه df بیشتر باشد، شکل نمودار به شکل توزیع استاندارد نزدیک و نزدیک تر می شود تا حدی که در df های زیاد(وقتی حجم نمونه زیاد است) توزیع t تقریبا عیناً شبیه توزیع z است.
- در حقیقت نمودار توزیع t وقتی df برابر با بی نهایت باشد، دقیقاً همان توزیع z است.


0.4
0.3
0.2
0.1
0.0
Probability
-4
-2
0
2
4
t
df = 1
df = 3
df = 8
df = 20
= standard normal

Student's t Distribution Table

For example, the t value for 18 degrees of freedom is 2.101 for 95% confidence interval (**2-Tail** $\alpha = 0.05$).



	90%	95%	97.5%	99%	99.5%	99.95%	1-Tail Confidence Level
	80%	90%	95%	98%	99%	99.9%	2-Tail Confidence Level
	0.100	0.050	0.025	0.010	0.005	0.0005	1-Tail Alpha
df	0.20	0.10	0.05	0.02	0.01	0.001	2-Tail Alpha
1	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192	
2	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991	
3	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240	
4	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103	
5	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688	
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588	
7	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079	
8	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413	
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809	
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869	
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370	
12	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178	
13	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208	
14	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405	
15	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728	
16	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150	
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651	
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216	
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834	
20	1.3253	1.7247	2.0860	2.5280	2.8453	3.8495	
21	1.3232	1.7207	2.0796	2.5176	2.8314	3.8193	
22	1.3212	1.7171	2.0739	2.5083	2.8188	3.7921	
23	1.3195	1.7139	2.0687	2.4999	2.8073	3.7676	
24	1.3178	1.7109	2.0639	2.4922	2.7969	3.7454	
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.7251	
26	1.3150	1.7056	2.0555	2.4786	2.7787	3.7066	
27	1.3137	1.7033	2.0518	2.4727	2.7707	3.6896	
28	1.3125	1.7011	2.0484	2.4671	2.7633	3.6739	
29	1.3114	1.6991	2.0452	2.4620	2.7564	3.6594	
30	1.3104	1.6973	2.0423	2.4573	2.7500	3.6460	

One sample t-test:

Individual	Angle
A	120.6
B	116.4
C	117.2
D	118.1
E	114.1
F	116.9
G	113.3
H	121.1
I	116.9
J	117.0

مثال: فرض کنید در حال تحقیق بر روی حس وضعی مفاصل انسان (توانایی ما برای تشخیص اینکه مفاصلمان در چه موقعیتی قرار دارد بدون اینکه آن ها را ببینید یا لمس کنید) هستید. شما می خواهید بدانید آیا مردم مفصل زانوی خود را بیشتر یا کمتر از حد واقعی تصور می کنند؟

شما ۱۰ فرد داوطلب را مورد آزمایش قرار می دهید و زانوی آن ها را تا ۱۲۰ درجه برای چند ثانیه خم می کنید. سپس زانو را به ۹۰ درجه برمی گردانید و از هر فرد می خواهید که زانو خود را مجدداً به ۱۲۰ درجه برگرداند.

زاویه ای که هر فرد به زانو خود می دهد را ثبت می کنید و به صورت زیر می شود.

الف- با اطمینان ۹۰٪ تعیین کنید که آیا مردم زاویه زانوی خود را در این شرایط متفاوت با حد مورد انتظارشان (۱۲۰ درجه) قرار می دهند؟

ب- با خطا ۲,۵٪ تعیین کنید که آیا مردم زاویه زانو خود را در این شرایط بیش از حد واقعی تصور می کنند؟

Paired samples t-test:

- وقتی می خواهیم میانگین دو دسته نمونه را باهم مقایسه کنیم و ببینیم آیا تفاوت معناداری بین جمعیت ها وجود دارد یا نه، اگر اعداد دو دسته نمونه ما یک به یک با هم جفت بوده و در واقع هر جفت عدد مربوط به یک نمونه باشد، از این تست استفاده می شود.
- در این تست دو متغیر را به یک متغیر حاصل از تفاضل آن ها تبدیل می کنیم. سپس برای یک دسته نمونه n عضوی مان one sample t-test انجام می دهیم.
- میانگین جمعیت فرضی مورد مقایسه صفر است.

Independent Two samples t-test:

گاهی اوقات لازم است تا دو دسته نمونه ای را با هم مقایسه کنیم که اعضا آن هیچ ربطی به هم ندارند.

وقتی مشاهده می کنیم که دو نمونه ما دارای میانگین متفاوت هستند به این شک می کنیم که آیا این دو نمونه از یک جمعیت گرفته شده اند یا این تفاوت بین دو نمونه معنادار است.

فرمول پیش بینی واریانس میانگین ها بر اساس s یک نمونه n عضوی برابر است با s^2/n.

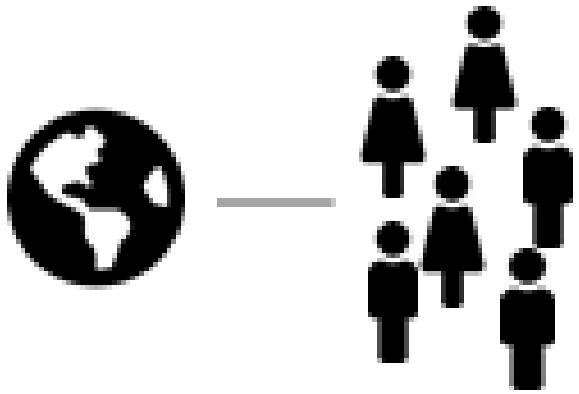
SEM اختلاف میانگین های دو نمونه n_1 و n_2 عضوی با انحراف معیار s_1 و s_2 از فرمول زیر قابل محاسبه است:

$$\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$$

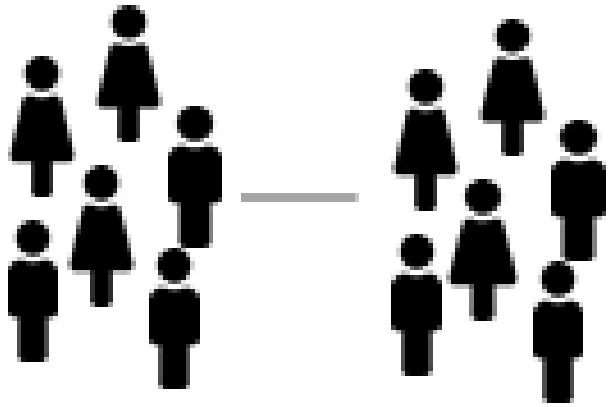
بدین ترتیب توزیعی مشخص برای تفاضل میانگین دو نمونه داریم که میانگین صفر و SEM برابر با $\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}$ دارد. بر همین اساس t-score ما برابر است با:

$$t = \frac{(x_1 - x_2)}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}}$$

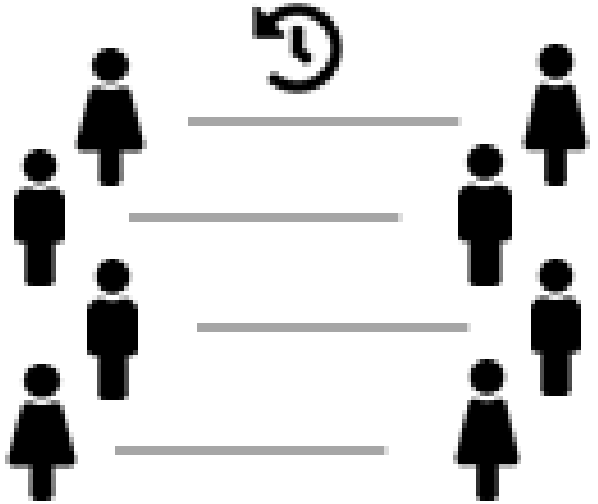
One Sample



Unpaired samples



Paired samples



آزمون مجذور کای:

به منظور آزمون فرضیه درباره استقلال فرآوانی ها که در طبقه های مختلف قرار گرفته اند به کار برده می شود.

پاسخ به این سوال که آیا بین فرآوانی ها مشاهده شده در نمونه با فرآوانی های فرضی اختلاف معناداری وجود دارد.

تست مجذور کای برای goodness of fit:

یکی از موارد استفاده این آزمون زمانی است که اطلاعات از یک نمونه جمع آوری شده باشد.

هرچه مجذور کای به دست آمده بزرگتر باشد، نشان دهنده آن است که بین اعداد مشاهده شده و اعداد موردنظر تفاوت معنادارتری وجود دارد.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

χ^2 = chi squared

O_i = observed value

E_i = expected value

توزیع مجذور کای:

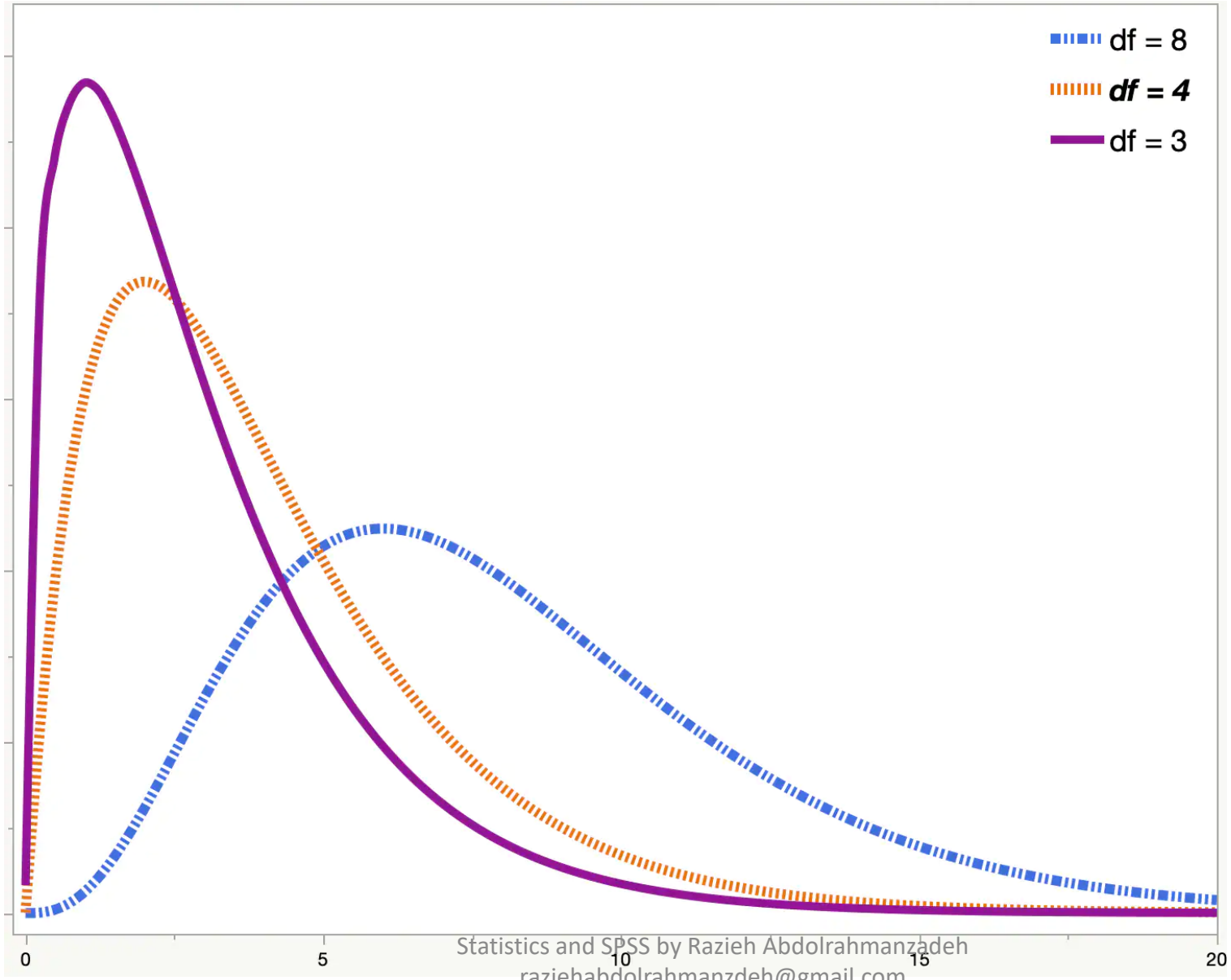
اگر برای تعداد زیادی واقعه که همگی در واقع از یک مدل آماری تبعیت می کنند، مجذور کای را حساب کنیم، این مجذور کای به احتمال بیشتری به صفر و به احتمال کمتری از صفر دور خواهد بود.

مختصات توزیع مجذور کای بر حسب اینکه درجه آزادی ما چقدر باشد، متفاوت است.

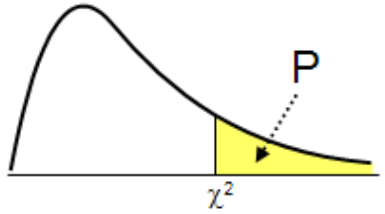
p-value برای یک آماره مجذور کای برابر است با احتمال رخداد آن مجذور کای و رخدادهای پرت تر از آن.

از آن جا مجذور کای همواره مثبت است، توزیع کای دو را همواره به صورت one tailed بررسی می کنیم.

توزیع مجذور کای برای برای درجه آزادی های مختلف:



جدول توزیع مجذور کای:



	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315
21	8.034	10.283	26.171	29.615	32.671	35.479	36.343	38.932	41.401	44.522	46.797
22	8.643	10.982	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.962	48.268
23	9.260	11.689	28.429	32.007	35.172	38.076	38.968	41.638	44.181	47.391	49.728
24	9.886	12.401	29.553	33.196	36.415	39.364	40.270	42.980	45.559	48.812	51.179
25	10.520	13.120	30.675	34.382	37.652	40.646	41.566	44.314	46.928	50.223	52.620
26	11.160	13.844	31.795	35.563	38.885	41.923	42.856	45.642	48.290	51.627	54.052
27	11.808	14.573	32.912	36.741	40.113	43.195	44.140	46.963	49.645	53.023	55.476
28	12.461	15.308	34.027	37.916	41.337	44.461	45.419	48.278	50.993	54.411	56.892
29	13.121	16.047	35.139	39.087	42.557	45.722	46.693	49.588	52.336	55.792	58.301
30	13.787	16.791	36.250	40.256	43.773	46.979	47.962	50.892	53.672	57.167	59.703
31	14.458	17.539	37.359	41.422	44.985	48.232	49.226	52.191	55.003	58.536	61.098

پایان

با تشکر از توجه شما

raziehabdolrahmanzadeh@gmail.com