Date:.................. Subject:..................

آزمون فرض (Hypothesis testing)

اگر مهندس تغذیه یک شرکت صنایع غذایی ، بخواهد بر مبنای داده‌های نمونه نظر دهد که مقدار اسید چرب ترانس موجود در محصول خود حداکثر 1,5 درصد است یا خیر، اگر یک کارشناس کشاورزی بخواهد بر مبنای آزمایش نظر دهد که آیا نوع خاص از کودهای کشاورزی محصول برنج بیشتری نسبت به کود دیگری تولید می‌کند یا خیر، یا یک سازنده دارو بر مبنای نمونه نظر دهد که آیا 90 درصد بیماران که داروی جدید را مصرف می‌کنند بهبود می‌یابند یا خیر ؛ همه این مسائل را می‌توان بزبان آزمون فرض آماری برگرداند.

در مورد اول و سوم ، شرکت سازنده باید p ، پارامتر یک جامعه دوجمله‌ای را معین نماید. در مورد دوم نیز کارشناس کشاورزی باید نظر دهد که آیا $\mu_1 > \mu_2$ که در آن μ_1 و μ_2 میانگین‌های دو جامعه نرمال می‌باشند.

مثال- برای کنترل نوعی سرطان روش استانداردی برای شیمی‌درمانی وجود دارد که 25 درصد روی بیماران تأثیر گذار است. محققان آزمایشگاهی روش دیگری را معرفی نموده‌اند و ادعا می‌کنند که اثر آن از روش استاندارد بیشتر است و حدود 50 درصد است. این آزمایشگاه 20 بیمار را بطور تصادفی مستقل از یکدیگر انتخاب می‌کند. متغیر تصادفی X را برابر با تعداد بیمارانی که بهبود یافته‌اند، قرار دهید. دراینصورت

$$X \sim B(20, p)$$

p پارامتر مجهول جامعه و درصد موثر بودن روش جدید است.

$$p = \frac{1}{2}$$

Date: .... Subject: ....

هرگاه بخواهیم یک ادعا را از طریق تأیید آن بوسیله اطلاعات حاصل از نمونه ثابت کنیم، خلاف آن ادعا را در فرض صفر (Null hypothesis) H_0 قرار می‌دهیم. خود ادعا را در فرض مقابل (Alternate hypothesis) H_1 قرار می‌دهیم.

$H_0: \quad p = \frac{1}{4}$

$H_1: \quad p = \frac{1}{2}$

اگر اطلاعات بدست آمده از نمونه با فرض H_0 سازگار باشد، گوییم فرض آماری رد (reject) شده است. این بدان معنی است که از روی نمونه انتخابی فرض آماری با قاطعیت رد شده است.

اگر اطلاعات بدست آمده از نمونه با فرض H_0 ناسازگار باشد، گوییم فرض آماری پذیرفته شده است (acceptance) و این بدان معناست که نمونه موجود دلیلی بر رد کردن آن فرض آماری را بدست نمی‌دهد.

آماره آزمون (statistic) مقدار محاسبه شده از یک نمونه، نامیده می‌شود.

نمونه تصادفی $(X_1, \dots, X_n) \rightsquigarrow T = T(X_1, \dots, X_n)$

مقادیر نمونه $(x_1, \dots, x_n)$ و $t = T(x_1, \dots, x_n)$ و

AZIN

Date:.................. Subject:..................

ناحیه بحرانی آزمون (critical region)

فرض کنید مجموعه C ، مجموعه مقادیری باشد که به ازای آنها آماره آزمون باعث رد فرض H_0 شود.

این مجموعه را ناحیه بحرانی آزمون می‌نامند.

مثال - در آزمون فرض $\begin{cases} H_0: & p=\frac{1}{4} \\ H_1: & p=\frac{1}{2} \end{cases}$ که در آن نمونه تصادفی $(X_1, \dots, X_n)$ را

انتخاب نمود، اگر متغیر تصادفی X ، تعداد افراد بهبود یافته باشد، آنگاه $X \sim B(20, p)$.

ناحیه قبول برای H_0 ، با مقادیر $x=0, 1, \dots, 6$ و ناحیه رد یا بحرانی متناظر

با مقادیر $x=7, 8, \dots, 20$ خواهد بود. پس هر یک از مجموعه‌های زیر را می‌توان به عنوان

ناحیه بحرانی در نظر گرفت:

$$C=\{x: \quad x \geq 7\}$$

$$C=\{x: \quad x \geq 8\} \quad \dots$$

خطاهای آزمون (Error)

۱- خطای نوع اول (Type I Error) : اگر H_0 درست باشد ولی ما آن را طبق نتیجه

حاصل از نمونه تصادفی رد می‌کنیم. احتمال خطای نوع اول عبارت است از:

$$\alpha = P(\text{رد } H_0 \mid \text{درست } H_0) = P(T(X_1, \dots, X_n) \in C \mid H_0)$$

به این خطا، سطح معنی‌دار یا سطح تشخیص آزمون می‌گویند.

Date: .... Subject: ....

2- خطای نوع دوم (Type II Error) : اگر H_0 نادرست باشد ولی ما طبق نتیجه بدست آمده از نمونه تصادفی آن را قبول می‌کنیم. احتمال خطای نوع دوم عبارت است از :

$$\beta = \mathbb{P}(H_0 \text{ قبول} \mid H_0 \text{ نادرست}) = \mathbb{P}(T(X_1, \dots, X_n) \notin C \mid H_1).$$

توان آزمون (power of a test)

توجه داشته باشید هدف از آزمون فرض، راهی برای اثبات ادعا H_1 است. پس اگر بخواهیم پارامتری برای توان یک آزمون فرض تعریف کنیم، می‌توان آن را بصورت زیر در نظر گرفت :

$$\beta^* = \mathbb{P}(H_0 \text{ رد} \mid H_0 \text{ نادرست}) = 1 - \mathbb{P}(H_0 \text{ قبول} \mid H_1) = 1 - \beta.$$

تذکر - در بین ناحیه‌های بحرانی یک آزمون، آن ناحیه که در آن α و β هر دو کوچک باشند مطلوب می‌باشد.

مثال - در مثال قبل ناحیه بحرانی $C = \{x : x \geq 9\}$ را در نظر بگیرید :

$$\alpha = \mathbb{P}(X \geq 9 \mid p = 1/4) = 1 - \mathbb{P}(X \leq 8 \mid p = 1/4) \underset{\text{طبق جدول}}{=} 1 - 0.9591 = 0.0409$$

$$\beta = \mathbb{P}(X \leq 8 \mid p = 1/2) \overset{\text{طبق جدول}}{=} 0.2517$$

مثال - حال ناحیه بحرانی $C = \{x : x \geq 10\}$ را در نظر بگیرید :

Date:................ Subject:................

$$C = \{ x : x \geq 10 \}$$

$$\alpha = P(X \geq 10 \mid p = \tfrac{1}{4}) = 1 - P(X \leq 9 \mid p = 1/4) \underset{\text{طبق جدول}}{=} 1 - 0.9861$$

$$\Rightarrow \alpha = 0.0139$$

$$\beta = P(X \leq 9 \mid p = 1/2) \underset{\text{طبق جدول}}{=} 0/4119$$

$$\beta^* = 1 - \beta = 0.5881$$

<u>تذکر</u>- تنها راهی که می‌توان احتمال‌های α و β هر دو را همزمان کم کرد، افزایش دادن اندازه نمونه n است.

<u>تذکر</u>- با تغییر ناحیه بحرانی، نمی‌توان هم خطای نوع اول و هم خطای نوع دوم را کاهش داد. پس باید از ناحیه‌های بحرانی ممکن، آن ناحیه‌ای را انتخاب کنیم که با اعمال یک کران بالا برای α، مقدار β را تا آنجا که می‌توان کاهش دهد، تا توان آزمون حداکثر شود!

<u>مثال</u>- فرض کنید $X \sim N(\mu, 4)$، آزمون فرض زیر را در نظر بگیرید:

$$\begin{cases} H_0 : \mu = 0 \\ H_1 : \mu = 1 \end{cases}$$

اگر $n = 25$ و $C = \{ (x_1, \dots, x_{25}) : \bar{x} > c \}$. مقدار c را طوری تعیین کنید که $\alpha = 0/1$. سپس مقادیر β و β^* را محاسبه نمایید.

Date:.................... Subject:....................

$$X \sim N(\mu, 4) \longrightarrow \frac{\bar{X} - \mu}{\frac{2}{5}} \sim N(0,1)$$

$$\alpha = P(\bar{X} > c \mid \mu = 0) = P\left(\frac{\bar{X} - \mu}{\frac{2}{5}} > \frac{c - 0}{\frac{2}{5}} \right)$$

$$= P\left(Z > \frac{5}{2}c\right) = 1 - P\left(Z \leq \frac{5}{2}c\right) = 0.1$$

$$\Rightarrow P\left(Z \leq 5/2\, c\right) = 0.9$$

$$\frac{5}{2}c = Z_{0.9} \underset{\text{طبق جدول}}{\approx} 1.28 \Longrightarrow c = \frac{1.28}{2.5} = 0.512$$

$$\beta = P(\bar{X} \leq 0.512 \mid \mu = 1) = P\left(\frac{\bar{X} - \mu}{\frac{2}{5}} \leq \frac{0.512 - 1}{\frac{2}{5}} \right) =$$

$$= P(Z \leq -1.22) = 0.1112$$

$$\beta^* = 1 - 0.1112 = 0.8888$$

انواع فرض

$$\begin{cases} H_0 : p = 1/4 \\ H_1 : p = 1/2 \end{cases}$$

فرض: ساده / مرکب

(Simple) ← تحت این فرض توزیع جامعه کاملاً مشخص می‌شود.

(composite) ← توزیع جامعه مشخص نمی‌شود!

$$\begin{cases} H_0 : p \leq 0.4 \\ H_1 : p > 0.4 \end{cases}$$

Date: .... Subject: ....

آزمون های فرض های آماری روی پارامترهای یک جامعه

الف_ فرض کنید X یک جامعه با میانگین μ و واریانس معلوم σ^2 است و $(X_1, \dots, X_n)$ یک نمونه تصادفی است و می‌خواهیم آزمون روی μ انجام دهیم. برای آن از برآوردگر $\bar{X}$ استفاده می‌کنیم.

(1)

$$\begin{cases} H_0 : & \mu = \mu_0 \\ H_1 : & \mu > \mu_0 \end{cases}$$

برآوردگر نقطه ای : $\bar{X}$

$$C = \{ (X_1, \dots, X_n) : \bar{X} > c \}$$

c را طوری تعیین می‌کنیم که $\alpha = \mathbb{P}(\bar{X} > c \mid \mu = \mu_0)$

در اینصورت اگر جامعه نرمال باشد یا نرمال نباشد ولی $n \geq 30$ ، داریم

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

برای تعیین c بصورت زیر عمل می‌کنیم:

$$\alpha = \mathbb{P}(\bar{X} > c \mid \mu = \mu_0)$$

$$= \mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{c - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right) = \mathbb{P}\left(Z > \frac{c - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right) \Rightarrow$$

$$1 - \alpha = \mathbb{P}\left(Z \leq \frac{c - \mu_0}{\sigma / \sqrt{n}}\right) \Rightarrow \frac{c - \mu_0}{\sigma/\sqrt{n}} = Z_{1-\alpha} \Rightarrow$$ طبق ناحیه بحرانی c می توان نتیجه گرفت

Date:.................................. Subject:...

نتیجه‌گیری : در آزمون $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ یا آزمون $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ فرض

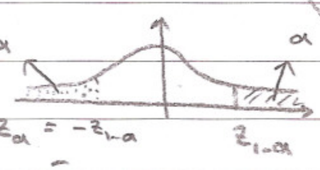
H_0 رد می‌گردد اگر و تنها اگر $Z = \dfrac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > Z_{1-\alpha}$

(2) $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases} \longrightarrow C = \{ (x_1, \dots, x_n) : \bar{X} < c \}$

حال طبق $Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ ، برای تعیین مقدار c در ناحیه بحرانی C از سطح معنی‌دار آزمون استفاده می‌کنیم.

$$\alpha = \mathbb{P}(\bar{X} < c \mid \mu = \mu_0) = \mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{c - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right) = \mathbb{P}\left(Z < \frac{c - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right)$$

$$\Rightarrow \frac{c - \mu_0}{\frac{\sigma}{\sqrt{n}}} = Z_\alpha = -Z_{1-\alpha}$$



طبق تعریف $Z_{1-\alpha}$ مساحت دم راست آن برابر α است و طبق متقارن بودن نمودار نرمال پس مساحت دم چپ نمودار نیز برابر α است.

پس طبق نحوه تعریف ناحیه بحرانی C می‌توان نتیجه گرفت :

در آزمون $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$ یا $\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$ ، فرض H_0 رد می‌گردد اگر و تنها اگر

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < -Z_{1-\alpha}.$$

Date: Subject:

(3)

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases} \Rightarrow C = \{(x_1, \dots, x_n) : \bar{X} < c_1 \vee \bar{X} > c_2\}$$

برای یافتن مقادیر c_1 و c_2 از سطح معنی‌دار آزمون استفاده می‌کنیم:

$$\alpha = P(\bar{X} < c_1 \vee \bar{X} > c_2 \mid \mu = \mu_0)$$

$$\Rightarrow 1-\alpha = P(\bar{X} \geq c_1 \wedge \bar{X} \leq c_2 \mid \mu = \mu_0)$$

$$= P(c_1 \leq \bar{X} \leq c_2 \mid \mu = \mu_0)$$

$$= P\left(\frac{c_1 - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq \frac{X - \mu}{\sigma/\sqrt{n}} \leq \frac{c_2 - \mu_0}{\sigma/\sqrt{n}}\right)$$

حال اگر مقادیر $\frac{c_1 - \mu_0}{\sigma/\sqrt{n}}$ و $\frac{c_2 - \mu_0}{\sigma/\sqrt{n}}$ را قرینه هم قرار دهیم، داریم:

$$Z_{1-\alpha/2} = \frac{c_2 - \mu_0}{\sigma/\sqrt{n}} = -\frac{c_1 - \mu_0}{\sigma/\sqrt{n}}$$

پس طبق ناحیه بحرانی C می‌توان نتیجه گرفت:

در آزمون $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$ فرض H_0 رد می‌شود اگر و تنها اگر

$$|Z| = \left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| > Z_{1-\alpha/2} .$$

Date:............................ Subject:............................

مثال- نسبت خانواده‌های یک شهر که از شرکت میهن شیر می‌خرند 0/6 است. اگر از یک نمونه تصادفی 10 خانواری 3 یا کمتر از شرکت میهن شیر بخرند، فرضیه صفر $p=0/6$ را به نفع فرضیه مقابل $p<0/6$ رد می‌کنیم. در این حالت سطح معنی‌دار آزمون را محاسبه نمائید. اگر $P=0/5$، احتمال خطای نوع دوم را محاسبه نمائید.

$X=$ تعداد خانواده‌هایی که از میهن شیر می‌خرند.　　$X \sim B(10, p)$

$$\begin{cases} H_0 : & p=0.6 \\ H_1 : & p<0/6 \end{cases}$$

$$C = \{ x : \ x \le 3 \}$$

$$\alpha = \mathbb{P}(X \le 3 \mid p=0.6) \overset{\text{طبق جدول}}{=} 0.0548$$

$$\beta = \mathbb{P}(X > 3 \mid p=0/5) = 1 - \mathbb{P}(X \le 3 \mid p=0.5)$$
$$= 1 - 0/1719 = 0/8281.$$

مثال- در صورتیکه از یک جامعه نرمال با واریانس 4 یک نمونه تصادفی به اندازه 16 انتخاب کرده باشیم، در آزمون

$$\begin{cases} H_0 : & \mu=8 \\ H_1 : & \mu=10 \end{cases}$$

Date:.............................. Subject:..

در سطح معنی‌دار $0/05$ ، احتمال خطای نوع دوم را می‌ سنجید (محاسبه کنید).

$X \sim N(\mu, 4)$ و $n=16$

$0/05 = P(\bar{X} > c \mid \mu = 8) = P\left(\frac{\bar{X}-\mu}{\frac{2}{4}} > \frac{c-8}{\frac{1}{2}}\right)$

$= P(Z > 2c-16) = 1 - P(Z \leq 2c-16)$

طبق جدول : $2c-16 = Z_{0/95} = 1.64 \Rightarrow c = 8.82$

$\Rightarrow \beta = P(\bar{X} \leq 8.82 \mid \mu = 10) = P\left(\frac{\bar{X}-\mu}{\frac{2}{4}} \leq \frac{8.82-10}{\frac{1}{2}}\right)$

$= P(Z \leq -2.35) = 0.0093$.

مثال - فرض کنید متغیر تصادفی X دارای تابع چگالی زیر باشد :

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0 \quad \theta > 0 \\ 0 & \text{else} \end{cases}$$

می‌خواهیم آزمون $\begin{cases} H_0 : \theta = 200 \\ H_1 : \theta = 500 \end{cases}$ را انجام دهیم. اگر $X > 300$ مشاهده شود،

فرض H_0 را رد می‌کنیم. احتمال خطای نوع اول و دوم و توان آزمون را محاسبه کنید.

Date:.............................. Subject:..............................

$$C = \{ x : x > 300 \}$$

$$\alpha = P(X > 300 \mid \theta = 200) = \int_{300}^{\infty} \frac{1}{200} e^{-\frac{x}{200}} dx$$

$$= -e^{-\frac{x}{200}} \Big|_{300}^{\infty} = e^{-\frac{3}{2}} = 0.2231$$

$$\beta = P(X \leq 300 \mid \theta = 500) = \int_{0}^{300} \frac{1}{500} e^{-\frac{x}{500}} dx$$

$$= -e^{-\frac{x}{500}} \Big|_{0}^{300} = 1 - e^{-\frac{3}{5}} = 0.4512$$

$$\beta^* = 1 - \beta = 0.5488$$

مثال - متوسط قد دانشجویان مرد سال اول یک دانشکده 68.5 اینچ با انحراف معیار 2.7 اینچ گزارش شده است. اگر یک نمونه تصادفی 50 تایی از دانشجویان سال اول نغلی دارای قد متوسط 69.7 اینچ باشد، آیا در سطح معنی‌دار 0.02 دلیلی بر تصور تغییر در صد متوسط قد وجود دارد؟

$$\begin{cases} H_0 : \mu = 68.5 \\ H_1 : \mu > 68.5 \end{cases}$$

$\alpha = 0.02$ و H_0 رد می‌شود اگر $\dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > Z_{1-\alpha}$

Date: .... Subject: ....

$$Z = \frac{69.7 - 68.5}{\frac{2.7}{\sqrt{5}}} = 3.143$$

$$Z_{1-\alpha} = Z_{0.98} = 2.05$$

$3.143 > 2.05 \longrightarrow$ پس H_0 رد می‌شود

یعنی تغییر در حد متوسط قد وجود دارد.

<u>مثال</u> ـ وزارت کار، حقوق روزانه ی کارگران کارخانه را بطور متوسط 132000 تومان با انحراف 25000 تعیین کرده است. اگر کارخانه‌ای به 40 کارگر خود روزانه بطور متوسط 122000 تومان پرداخت نماید، آیا می‌توان این کارخانه را متهم نمود که کمتر از میزان تعیین شده وزارت کار، پرداخت می‌کند؟

آزمون فرض
$$\begin{cases} H_0: \mu = 132000 \\ H_1: \mu < 132000 \end{cases}$$

در سطح معنی دار α، فرض H_0 درصورتی که $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} < -Z_{1-\alpha}$ باشد، رد می‌شود:

$$Z = \frac{122000 - 132000}{\frac{25000}{\sqrt{40}}} = -2.53 < -Z_{1-\alpha}$$

طبق جدول

$= Z_{0.994}$

اگر در مساله مقدار α مشخص نشده بود از مقدار $\alpha = 0.05$ استفاده می‌کنیم

اگر سطح معنی دار $\alpha = 0.05$ در نظر بگیریم $Z_{0.95} = 1.64$ پس می‌توان کارخانه را متهم کرد.

Date:.................. Subject:..................

مثال- طول فنرهایی که یک شرکت تولید می کند دارای توزیع نرمال با انحراف معیار 6 سانتی متر است. سازنده فنرها ادعا می کند که متوسط طول فنرها بیش از 65 سانتی متر است. خریداری مایل به آزمون ادعای سازنده است. از نظر خریدار خرید فنرهایی که متوسط طول آنها حداکثر 65 سانتی متر است مقرون به صرفه نیست. بدین منظور نمونه‌ای مرکب از 9 فنر را انتخاب می کند و طول آنها را اندازه‌گیری می کند.

الف- در مورد میانگین طول فنرها، فرض‌ها را فرمول بندی و ناحیه بحرانی آزمون را در سطح معنی‌دار 0/05 تعیین کنید.

ب- اگر $\mu = 67$، احتمال خطای نوع دوم و توان آزمون را محاسبه نمایید.

ج- اگر نتایج حاصل از بررسی 9 فنر به قرار زیر باشد، خریدار چه تصمیمی می گیرد؟

65 , 64 , 66 , 65 , 59 , 69 , 64 , 72 , 63

پاسخ - الف -

طول فنر $X \sim N(\mu, 36)$

آزمون فرض
$$\begin{cases} H_0: & \mu = 65 \\ H_1: & \mu > 65 \end{cases}$$

$n = 9 \quad , \quad \alpha = 0/05$

$$C = \{(X_1, \dots, X_9) : \bar{X} > \gamma\}$$

$$\alpha = P(\bar{X} > \gamma \mid \mu = 65) = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{\gamma - 65}{\frac{6}{3}}\right) = 1 - P\left(Z \le \frac{\gamma - 65}{2}\right)$$

$$\frac{\gamma - 65}{2} = Z_{1-\alpha} = Z_{0.95} = 1.65 \Rightarrow \gamma = 2 \times 1.65 + 65 = 68.3$$

$\mu > 65$ و $\mu = 67$

Date: .... Subject: ....

ب -
$$\beta = \mathbb{P}(\bar{X} \leq 63.3 \mid H_1) = \mathbb{P}(\bar{X} \leq 63.3 \mid \mu = 67)$$

$$= \mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{63.3 - 67}{6/3} \right) = \mathbb{P}(Z \leq 0.645) = 0.742$$

$$\beta^* = 1 - \beta = 0.26$$

ج -

$$\bar{x} = \frac{587}{9} = 65.22 < \gamma = 68.3$$

پس دلیلی به رد فرض H_0 نیست. پس خریدار از خریدن [illegible] منصرف می‌شود.

Date:.................. Subject:..................

مثال – ۷ سب‌های تلویزیون موجود در بازار دارای میانگین طول عمر 1200 ساعت با انحراف معیار 300 ساعت هستند. کارخانه‌ای ۷ سب جدیدی تولید کرده و مدعی است که میانگین طول عمر ۷ سب‌های ساخت اش بیشتر از 1200 ساعت است. برای بررسی این ادعا یک نمونه 100 تایی انتخاب و میانگین طول عمر 1265 ساعت بدست آمده است. در سطح معنی‌دار ۰/۰۱ آیا ادعا کارخانه درست است.

$$\begin{cases} H_0: \mu = 1200 \\ H_1: \mu > 1200 \end{cases}, \quad \alpha = 0/01$$

$Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} > Z_{1-\alpha} = Z_{0/99} = 2.33$ اگر تنها اگر داشته باشیم H_0 رد می‌شود

$$Z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{1265-1200}{\frac{300}{\sqrt{100}}} = 2.167 > 2.33 \quad \times$$

پس H_0 رد نمی‌شود و ادعای کارخانه مورد قبول قرار نمی‌گیرد (البته در سطح معنی‌دار ۰/۰۱).

مثال – یک فرآیند تولید ذغال سنگ موجود است که توزیع تولید روزانه آن نرمال با میانگین 800 و انحراف معیار 30 تن است. بمنظور ازدیاد تولید اصلاحی در این فرآیند پیشنهاد شده است و یک نمونه تصادفی 100 روزه از تولید فرآیند اصلاح شده دارای میانگین 812 تن است. در سطح معنی‌دار ۰/۰۱ آیا فرآیند اصلاح شده میانگین تولید روزانه را افزایش می‌دهد؟

$$\begin{cases} H_0: \mu = 800 \\ H_1: \mu > 800 \end{cases} \qquad \alpha = 0/01 \rightarrow Z_{1-\alpha} = Z_{0/99} = 2.33$$

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{812-800}{\frac{30}{\sqrt{100}}} = 4 > 2.33 \checkmark$$

پس H_0 رد می‌گردد یعنی فرآیند اصلاح شده میانگین روزانه را افزایش می‌دهد.

Date:.................... Subject:....................

ب ـ آزمون فرض روی میانگین با نامعلوم بودن واریانس . طبق مطلب معرفی قبل اگر جامعه نرمال باشد.

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}$$

انواع آزمون فرض برای این قسمت عبارتست از :

① $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \rightarrow C = \{ (X_1, \dots, X_n) : \bar{X} > c \}$

$$\alpha = \mathbb{P}(\bar{X} > c \mid \mu = \mu_0) = \mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} > \frac{c - \mu_0}{\frac{S}{\sqrt{n}}} \right) = \mathbb{P}\left(T > \frac{c - \mu_0}{\frac{S}{\sqrt{n}}} \right)$$

$$\mathbb{P}\left(T \leq \frac{c - \mu_0}{\frac{S}{\sqrt{n}}} \right) = 1 - \alpha \xrightarrow{\frac{c-\mu_0}{S/\sqrt{n}} = t_{1-\alpha}(n-1)}$$

پس طبق ناحیه بحرانی داریم :

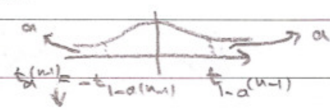
نتیجه: فرض H_0 رد می‌گردد اگر و تنها اگر

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} > t_{1-\alpha}(n-1) .$$

② $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \rightarrow C = \{ (X_1, \dots, X_n) : \bar{X} < c \}$

$$\alpha = \mathbb{P}(\bar{X} < c \mid \mu = \mu_0) = \mathbb{P}\left(\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < \frac{c - \mu_0}{\frac{S}{\sqrt{n}}} \right) = \mathbb{P}\left(T < \frac{c - \mu_0}{\frac{S}{\sqrt{n}}} \right)$$

$$\rightarrow \frac{c - \mu_0}{\frac{S}{\sqrt{n}}} = t_{\alpha}(n-1) = -t_{1-\alpha}(n-1)$$



فرض H_0 رد می‌گردد اگر و تنها اگر

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < -t_{1-\alpha}(n-1) .$$

AZIN

Date:............ Subject:............

(3) $\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases}$ $\quad C = \{(x_1, \dots, x_n) : \bar{X} < c_1 \vee \bar{X} > c_2\}$

$$\alpha = \mathbb{P}(\bar{X} < c_1 \vee \bar{X} > c_2 \mid \mu = \mu_0) = 1 - \mathbb{P}(\bar{X} \geq c_1 \wedge \bar{X} \leq c_2 \mid \mu = \mu_0)$$

$$= 1 - \mathbb{P}\left(\frac{c_1 - \mu_0}{\frac{s}{\sqrt{n}}} \leq \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \leq \frac{c_2 - \mu_0}{\frac{s}{\sqrt{n}}}\right)$$

$$\Rightarrow \mathbb{P}\left(\frac{c_1 - \mu_0}{\frac{s}{\sqrt{n}}} \leq T \leq \frac{c_2 - \mu_0}{\frac{s}{\sqrt{n}}}\right) = 1 - \alpha$$

حال اگر قرار دهیم:

$$t_{1-\alpha/2}(n-1) = \frac{c_2 - \mu_0}{\frac{s}{\sqrt{n}}} = -\frac{c_1 - \mu_0}{\frac{s}{\sqrt{n}}}.$$

نتیجه مطلوب حاصل می‌گردد. پس طبق ناحیه بحرانی می‌توان نتیجه گرفت:

فرض H_0 رد می‌گردد اگر و تنها اگر $|T| = \left|\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}\right| > t_{1-\alpha/2}(n-1).$

Date:.................... Subject:....................

مثال- نمرات امتحان هوش یک دبیرستان دارای توزیع نرمال با انحراف معیار 6 است. اگر یک رده 9 نفری

امتحان گرفته ایم و نتایج زیر بدست آمده است: 63 و 72 و 64 و 69 و 59 و 65 و 66 و 64 و 65

مدیر دبیرستان ادعا می کند که میانگین نمرات کل دبیرستان از 65 بیشتر است. در سطح معنی دار 0.05 آیا ادعای

مدیر دبیرستان را قبول می کنید؟

$$\begin{cases} H_0: & \mu \leq 65 \\ H_1: & \mu > 65 \end{cases}$$

$n = 9$ و σ^2 نامعلوم

H_0 رد می گردد اگر و تنها اگر

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} > t_{0/95}(8) = 1.86$$

طبق جدول

$$S^2 = \frac{1}{8}\left[\sum x_i^2 - \frac{1}{9}\left(\sum x_i\right)^2\right] = \frac{1}{8}\left(38363 - \frac{1}{9}\;587^2\right] = 13.44$$

$$\bar{X} = \frac{587}{2} = 65.22$$

$$\Rightarrow \quad T = \frac{65.22 - 65}{\sqrt{\frac{13.44}{9}}} = 0/18 < 1.86 \rightarrow$$ H_0 رد نمی گردد و ادعای مدیر را

رد می کنیم.

▲

مثال- مطالعه ی آماری در گذشته نشان داده است که میزان غیبت کارمندان بخاطر بیماری در عرض یک سال دارای توزیع نرمال

با میانگین 15 روز است. پژوهشگری برای 25 کارمند در سال گذشته تعداد روزهای غیبت را به شرح زیر ثبت کرده:

5, 25, 10, 0, 3, 50, 12, 14, 40, 12, 32, 5, 8, 4, 47, 20, 14, 18, 16

10, 1, 22, 58, 23, 9

در سطح معنی دار 0/05 بررسی نمائید که در سال گذشته متوسط روزهای غیبت بیش از 15 روز است.

Date:.............................. Subject:..............................

$$\begin{cases} H_0: \mu = 15 \\ H_1: \mu > 15 \end{cases}$$

$n=25$ ، σ^2 نامعلوم

H_0 را رد می‌کنیم اگر و تنها اگر

$$T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} > t_{0,95}(24) = 1.71$$

$$\bar{X} = \frac{458}{25} = 18.32$$

$$S^2 = \frac{1}{24}\left[\sum x_i^2 - \frac{1}{25}\left(\sum x_i\right)^2\right] = \frac{1}{24}\left[14416 - \frac{1}{25}\,458^2\right] = 251.06$$

$$T = \frac{18.32 - 15}{\sqrt{\frac{251.06}{25}}} = 1.048 < 1.71$$

پس دلیلی بر رد H_0 نداریم و میانگین تعداد غیبت 15 روز است.

مثال- یک کارخانه مواد شیمیایی بگونه‌ای طراحی شده است که روزانه بطور متوسط 800 تن محصول داشته باشد.

در 5 روز متوالی، این کارخانه به ترتیب 785 ، 805 ، 790 ، 793 ، 802 تن

محصول داشته است. با فرض نرمال بودن میزان محصول، آیا این داده‌ها نشان دهنده کاهش در حد متوسط میزان محصول کارخانه است؟

$$\begin{cases} H_0: \mu = 800 \\ H_1: \mu < 800 \end{cases}$$

σ^2 نامعلوم ، $n=5$ و سطح معنی‌دار داده شده $\alpha = 0/05$

H_0 را رد می‌کنیم اگر و تنها اگر

$$T = \frac{\bar{x}-\mu}{\frac{S}{\sqrt{n}}} < -t_{0.95}(4) = -2.13$$

$$\bar{X} = \frac{3975}{5} = 795$$

$$S^2 = \frac{1}{4}\left[3160403 - \frac{1}{5}(795)^2\right] = 69.5$$

AZIN

$$T = \frac{795 - 800}{\sqrt{\frac{69.5}{5}}} = -1.341 > -2.13$$

پس H_0 رد نمی‌گردد.

Date:............................ Subject:...

مثال- وزن یک نمونه تصادفی ۱۰ تایی از حلب های روغن برحسب کیلوگرم بصورت زیر گزارش شده است:

10.1 و 10.3 و 9.7 و 10.1 و 9.8 و 10.2 و 10.3 و 10.4 و 9.8 و 9.9

با فرض نرمال بودن وزن حلب های روغن، آیا در سطح معنی دار ۰/۰۱ می توان متقاعد شد که میانگین وزن این حلب های روغن ۱۰ کیلوگرم است؟

$$\begin{cases} H_0: & \mu = 10 \\ H_1: & \mu \neq 10 \end{cases}$$

σ^2 نامعلوم - $n=10$ و $\alpha = 0/01$

H_0 رد می شود اگر و تنها اگر

$$|T| = \left| \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \right| > t_{0.995}(9) = 3.25$$

طبق جدول

$$\bar{x} = \frac{100.6}{10} = 10.06$$

$$s^2 = \frac{1}{9}\left[1012.58 - \frac{1}{10}\, 100.6^2\right] \doteq 0.0604$$

$$T = \frac{10.06 - 10}{\sqrt{\frac{0/0604}{10}}} = 0.771 < 3.25 \longrightarrow$$ H_0 رد نمی شود و میانگین ۱۰ کیلو است.

مثال- طبق یک تئوری فیزیکی، میانگین افزایش درجه حرارت آبی که به عنوان خنک کننده در یک اتاقک کمپرسور بکار می رود، نباید بیشتر از 5 درجه سانتیگراد باشد. افزایش درجه حرارت در 8 بار استفاده مستقل از کمپرسور، اندازه گیری شده است: 5.1 و 6.4 و 5.9 و 6.5 و 4.9 و 5.7 و 6.3 و 6.4

اگر افزایش درجه حرارت دارای توزیع نرمال باشد، آیا این داده ها با تئوری فیزیکی متناقض هستند؟

Date:.................... Subject:....................

$$\begin{cases} H_0 = & \mu \leq 5 \\ H_1 : & \mu > 5 \end{cases}$$

σ^2 نامعلوم — $n = 8$ و α داده نشده $\rightarrow \alpha = 0.05$

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} > t_{0.95}(7) \underset{\text{طبق جدول}}{=} 1.9$$

H_0 رد می‌گردد اگر داشته باشیم

$$\bar{X} = \frac{47.2}{8} = 5.9$$

$$S^2 = \frac{1}{7}\left[281.18 - \frac{1}{8} 47.2^2\right] = 0.3857$$

$$T = \frac{5.9 - 5}{\sqrt{\frac{0.3857}{8}}} = 4.099 > 1.9 \rightarrow$$

H_0 رد می‌گردد یعنی میانگین افزایش دما با آنتن مشاهده شده است.

AZIN

Date:.................... Subject:....................

ج – آزمون فرض برای واریانس جامعه : طبق مطالب مطرح شده در فصل قبل اگر جامعه نرمال باشد،

$$Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

انواع آزمون فرض برای این قسمت عبارتست از :

① $\begin{cases} H_0: & \sigma^2 = \sigma_0^2 \\ H_1: & \sigma^2 > \sigma_0^2 \end{cases} \rightarrow C = \{(x_1, \dots, x_n) : S^2 > c\}$

برای بدست آوردن c از سطح معنی دار استفاده می‌کنیم،

$$\alpha = P(S^2 > c \mid \sigma^2 = \sigma_0^2) = P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)c}{\sigma_0^2}\right) = P\left(Y > \frac{(n-1)c}{\sigma_0^2}\right)$$

$$1-\alpha = P\left(Y \le \frac{(n-1)c}{\sigma_0^2}\right) \Rightarrow \chi^2_{1-\alpha}(n-1) = \frac{(n-1)c}{\sigma_0^2}$$

درنتیجه طبق ناحیه بحرانی می‌توان گفت :

فرض H_0 رد می‌گردد اگر و تنها اگر $Y = \frac{(n-1)S^2}{\sigma^2} > \chi^2_{1-\alpha}(n-1)$

② $\begin{cases} H_0: & \sigma^2 = \sigma_0^2 \\ H_1: & \sigma^2 < \sigma_0^2 \end{cases} \rightarrow C = \{(x_1, \dots, x_n) : S^2 < c\}$

$$\alpha = P(S^2 < c \mid \sigma^2 = \sigma_0^2) = P\left(\frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)c}{\sigma_0^2}\right) \Rightarrow \chi^2_{\alpha}(n-1) = \frac{(n-1)c}{\sigma_0^2}$$

حال طبق ناحیه بحرانی می‌توان گفت :

فرض H_0 رد می‌گردد اگر و تنها اگر $Y = \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha}(n-1)$

Date: .... Subject: ....

3) $\begin{cases} H_0: & \sigma^2 = \sigma_0^2 \\ H_1: & \sigma^2 \neq \sigma_0^2 \end{cases} \Rightarrow C = \{ (X_1, \dots, X_n) : \; S^2 > c_1 \;\text{یا}\; S^2 < c_2 \}$

$$\alpha = \mathbb{P}(S^2 > c_1 \vee S^2 < c_2 \mid \sigma^2 = \sigma_0^2) = 1 - \mathbb{P}(c_2 \leq S^2 \leq c_1 \mid \sigma^2 = \sigma_0^2)$$

$$= 1 - \mathbb{P}\left(\frac{(n-1)c_2}{\sigma_0^2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \frac{(n-1)c_1}{\sigma_0^2} \right)$$

$$\Rightarrow \mathbb{P}\left(\frac{(n-1)c_2}{\sigma_0^2} \leq Y \leq \frac{(n-1)c_1}{\sigma_0^2} \right) = 1 - \alpha$$

پس می‌توان طبق ناحیه بحرانی گفت

فرض H_0 رد می‌شود اگر و تنها اگر

$$\begin{cases} Y < \chi^2_{\alpha/2}(n-1) \\ \text{یا} \quad Y > \chi^2_{1-\alpha/2}(n-1) \end{cases}$$

Date:.................................. Subject:..

مثال- متوسط درجه حرارت سالانه یک شهر از میانگین متوسط درجه حرارت پانزدهمین روز هر ماه سال اندازه‌گیری شده است. انحراف معیار درجه حرارت سالانه شهری در یک دوره ۱۰۰ ساله 16 درجه فارنهایت بوده است. در مدت 15 سال گذشته انحراف استاندارد درجه حرارت سالانه را محاسبه نموده‌اند که برابر 18 درجه فارنهایت بوده است. آیا این اطلاعات دلیلی بر این است که انحراف معیار درجه حرارت سالیانه این شهر از انحراف معیار درجه حرارت سابق شهر بیشتر است؟

$$\begin{cases} H_0: & \sigma = 16 \\ H_1: & \sigma > 16 \end{cases} \qquad n = 15, \quad s = 18$$

فرض H_0 رد می‌گردد اگر و تنها اگر

$$Y = \frac{(n-1)S^2}{\sigma^2} > \chi^2_{0.95}(14) = 23.7$$

$$Y = \frac{14 \times 18^2}{16^2} = 17.719 < 23.7$$

پس دلیلی برای رد H_0 نمی‌باشد.

مثال- محققی معتقد است که واریانس اندازه‌هایی که در طول آزمایش ثبت می‌کند کوچک‌تر از 2 است. در یک آزمایش، او اندازه‌های 4.1، 5.2 و 10.2 را ثبت کرده است. اگر اندازه‌ها دارای توزیع نرمال باشند آیا می‌توان ادعای محقق را در سطح معنی‌دار ۰/۰۱ پذیرفت؟

$$\begin{cases} H_0: & \sigma^2 \geq 2 \\ H_1: & \sigma^2 < 2 \end{cases} \qquad \alpha = 0/01$$

H_0 رد می‌گردد اگر و تنها اگر

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

$$\bar{x} = \frac{19.5}{3}$$

$$s^2 = \frac{1}{2}\left(147.89 - \frac{1}{3}\, 19.5^2\right) = 10.57$$

Date: .... Subject: ....

$$Y = \frac{2 \times 10.57}{2} = 10.57 > \chi^2_{0.01}(2) = 0.02$$

H_0 رد نمی‌گردد.

مثال: قطعه زمینی بوسیله 5 دانشجوی نقشه برداری، مساحی شده و مساحت‌های زیر برحسب جریب بدست آمده اند 7.23 , 7.28 , 7.21 , 7.24 , 7.27

با فرض نرمال بودن اندازه‌گیری، در سطح معنی‌دار 0.05 آیا می‌توان ادعا کرد که واریانس 0.9 است؟

$$\begin{cases} H_0: & \sigma^2 = 0.9 \\ H_1: & \sigma^2 \neq 0.9 \end{cases}$$

$$s^2 = \frac{1}{4}\left(262.5259 - \frac{1}{5}\,36.23^2\right) = 0.00083$$

$$\chi^2_{0.025}(4) = 0.484$$

$$\chi^2_{0.975}(4) = 11.1$$

$$Y = \frac{4 \times 0.00083}{0.9} = 0.00369 < 0.484$$

پس H_0 رد می‌گردد. یعنی واریانس اندازه‌گیری 0.9 نیست.

AZIN

آزمون برازش

تا الآن فرض می‌کردیم جامعه دارای توزیع احتمال خاص است، حال در مورد پارامترهای مجهول آن یعنی μ یا σ^2 آزمون را انجام می‌دادیم. اما اگر توزیع احتمال خود جامعه نیز مجهول باشد، از آزمون برازش استفاده می‌کنیم. به جامعه توزیعی را نسبت می‌دهیم و سپس از طریق آزمون مربع-کای، آن توزیع را ارزیابی می‌کنیم.

فرض کنید یک آزمایش تصادفی دارای نتایج متمایز $c_1, \dots, c_k$ است. احتمال وقوع هر یک از آنها p_i $(i=1,\dots,k)$ می‌باشد بطوری که $\sum_{i=1}^{k} p_i = 1$.

این آزمایش را n بار بطور مستقل تکرار می‌کنیم. اگر

$O_i =$ تعداد دفعاتی که نتیجه آزمایش برابر c_i آید

بوضوح $\sum_{i=1}^{k} O_i = n$. $\quad O_i \sim B(n, p_i) \quad i=1,\dots,k;$

آزمون برازش:

$$\begin{cases} H_0: & p_1 = t_1, \dots, p_k = t_k \\ H_1: & \text{حداقل یکی از } p_i \neq t_i \text{ باشد.} \end{cases}$$

که در این آزمون t_i ها مقادیری مشخص می‌باشند که $\sum t_i = 1$ و $0 < t_i < 1$.

اگر فرض H_0 درست باشد، $E(O_i) = nt_i$، این مقدار را برابر e_i قرار دهید.

$$\forall i \quad e_i = nt_i \Rightarrow \sum_{i=1}^{k} e_i = n$$

Date:.............. Subject:..............

حال از آماره پیرسون

$$Y = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

استفاده می‌کنیم. اگر $o_i \geq 5$ ، بطور تقریبی $Y \sim \chi^2_{(k-1)}$.

پس اگر فرض H_0 درست باشد ، اختلاف e_i و o_i کم است پس Y خیلی کوچک است.

پس ناحیه بحرانی آزمون عبارت است از :

$$C = \left\{ Y : \quad Y = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} > r \right\}$$

که مقدار r را با توجه به سطح معنادار آزمون یعنی α تعیین می‌توان نمود.

نتیجه‌گیری - در آزمون برازش ، فرض H_0 رد می‌گردد اگر و تنها اگر

$$Y = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} > \chi^2_{1-\alpha}(k-1) .$$

تذکر - در آزمون برازش اگر $o_i < 5$ بود ، باید برخی از نتایج آزمایش تصادفی را با یکدیگر ادغام کنیم تا این شرط حاصل گردد.

تذکر - اگر e_i ها به پارامترهای مجهول دیگری وابسته بودند ، ابتدا آن پارامترها را برآورد می‌کنیم ، سپس به ازای هر یک از آنها پارامترها از درجه آزادی $\chi^2_{1-\alpha}$ می‌کاهیم.

Date: .... Subject: ....

مثال - داده‌های زیر میزان محصول ذرت را در ۱۰۰ مزرعه نشان می‌دهد. آیا در این مزارع

$\sum x_i = 91400$ ، $S = 331.8$ ، آیا میزان محصول ذرت این مزارع از توزیع نرمال پیروی می‌کند؟

	محصول برحسب کیلوگرم	تعداد مزارع
1	$99.5 \le X < 299.5$	3
2	$299.5 \le X < 499.5$	7
3	$499.5 \le X < 699.5$	15
4	$699.5 \le X < 899.5$	26
5	$899.5 \le X < 1099.5$	22
6	$1099.5 \le X < 1299.5$	13
7	$1299 \le X < 1499.5$	9
8	$1499 \le X < 1699.5$	5

حل - X = میزان محصول ذرت

$$\begin{cases} H_0 : X \sim N(\mu, \sigma^2) \\ H_1 : X \nsim N(\mu, \sigma^2) \end{cases}$$

$$\mu = \bar{X} = \frac{91400}{100} = 914$$

$$\sigma^2 = S^2 = (331.8)^2$$

$$e_0 = P(X < 99.5) = P\left(Z < \frac{99.5 - 914}{331.8}\right) = P(Z < -2.45)$$

$= 0/0071$

AZIN

$e_0 = 0/71$

Date:............................ Subject:..

$$t_1 = \mathbb{P}(99.5 \leq X < 299.5) = \mathbb{P}\left(\frac{99.5-914}{331.8} \leq Z < \frac{299.5-914}{331.4}\right)$$

$$= \mathbb{P}(-2.45 \leq Z < -1.85) = 0.0322 - 0.0071 = 0.0251$$

$$e_1 = 2.51$$

$$t_2 = \mathbb{P}(299.5 \leq X < 499.5) = \mathbb{P}(-1.85 \leq Z < -1.25)$$

$$= 0.1056 - 0.0322 = 0.0734$$

$e_2 = 7.34$

$e_3 = 15.22$

$e_4 = 22.62$

$e_5 = 22.83$

$e_6 = 16.47$

$e_7 = 8.38$

$e_8 = 3.03$

$e_9 = 0.89$

i	شماره کمتر از 3	3	4	5	6	شماره بیشتر از 6
o_i	10	15	26	22	13	14
e_i	10.56	15.22	22.62	22.83	16.47	12.3

$$Y = \sum_{i=1}^{6} \frac{(o_i - e_i)^2}{e_i} = 1.53$$

AZIN

$$\chi^2_{0.95}(6-1-\underbrace{2}_{\hat{\mu}\,\hat{\sigma}^2}) = \chi^2_{0.95}(3) = 7.81 \Rightarrow$$ H_0 رد نمی‌گردد

Date:............................ Subject:..

مثال- مشاهدات زیر تعداد قطعات خراب در 150 کارتن محتوی قطعات تولید شده توسط یک کارخانه را نشان می‌دهد:

تعداد قطعات خراب	0	1	2	3	4	5	6	7
تعداد کارتن‌ها	23	39	43	23	10	7	4	1

آیا تعداد قطعات خراب در کارتن‌های این کارخانه از توزیع پواسون پیروی می‌کند؟

تابع احتمال توزیع پواسون $\quad X \sim P(\mu) \Longrightarrow f_X(x) = \frac{e^{-\mu}\mu^{x}}{x!} \quad x=0,1,\dots$

$\mu = E[X] = Var[X]$

$X =$ تعداد قطعات خراب در هر کارتن

$$\begin{cases} H_0: X \sim P(\mu) \\ H_1: X \nsim P(\mu) \end{cases}$$

برآورد μ : $\quad \bar{X} = \frac{\sum x_i \times \text{تعداد تکرار هر نمونه}}{n} = \frac{1}{150}\sum_{i=0}^{7} n_i \cdot x_i$

$$= \frac{1}{150}(0\times 23 + 1\times 39 + 2\times 43 + \dots + 7\times 1) = 2$$

$$\Rightarrow \forall x = 0,1,\dots \qquad t_i = f_X(x=i) = \frac{e^{-2}\,2^{i}}{i!}$$

i	0	1	2	3	4	5	6	7
o_i	23	39	43	23	10	7	4	1
t_i	0/1353	0/2706						0/0045
e_i	20.3	40.6	40.6	27.06	13.53	5.41	1.8	0.7

Date: .... Subject: ....

چون مقادیر $o_6<5$ و $o_7<5$ پس این دو تا را با هم ادغام می‌کنیم ،

i	0	1	2	3	4	5	بزرگتر از 5
o_i	23	39	43	23	10	7	5
e_i	20.3	40.6	40.6	27.06	13.53	5.41	2.5

$$Y=\sum_{i=0}^{6}\frac{(o_i-e_i)^2}{e_i}=\frac{(20.3-23)^2}{20.3}+\cdots+\frac{(2.5-5)^2}{2.5}$$

$$=0.36+0.06+0.14+0.61+0.92+0.47+2.5$$

$$=5.06$$

$\rightarrow k=7$

p_i ها به برآورد پارامتر λ وابسته

$$\alpha=0.05\longrightarrow \chi^2_{1-\alpha}(k-1-1)=\chi^2_{.95}(5)=11.1$$

H_0 رد می‌گردد اگر

$Y=5.06>11.1$

×

پس H_0 رد نمی‌گردد و توزیع پوآسون بر داده‌ها برازنده است !

AZIN

Date:........................ Subject:........................

مثال- اگر S_1^2 و S_2^2 به ترتیب واریانس‌های دو نمونه تصادفی مستقل از دو جامعه نرمال باشد و بدانیم

که واریانس جامعه دوم سه برابر واریانس جامعه اول است. و نمونه‌ها به اندازه $n_1 = 8$ و $n_2 = 12$

انتخاب شده باشند، مطلوبست محاسبه $\mathbb{P}(S_1 < \sqrt{1.63}\, S_2)$

$$\mathbb{P}(S_1 < \sqrt{1.63}\, S_2) = \mathbb{P}(S_1^2 < 1.63\, S_2^2)$$

$$= \mathbb{P}\left(\frac{S_1^2}{S_2^2} < 1.63\right)$$

$$= \mathbb{P}\left(\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} < \frac{\sigma_2^2}{\sigma_1^2}\, 1.63\right)$$

$$= \mathbb{P}(\underset{F \sim F(7, 11)}{F} < \underbrace{3 \times 1.63}_{4.89}) = 0.99$$

مثال- میانگین و انحراف معیار نمرات درس آمار دانشجویان به ترتیب 72 و 8 است.

دو نمونه تصادفی مستقل 36 و 40 تایی انتخاب می‌کنیم، احتمال اینکه تفاضل میانگین نمرات

این دو نمونه بین 2 و 5 باشد را بیابید.

$$\bar{X}_1 - \bar{X}_2 \sim N\left(72 - 72, \frac{8^2}{36} + \frac{8^2}{40}\right) = N(0, 3.4)$$

AZIN

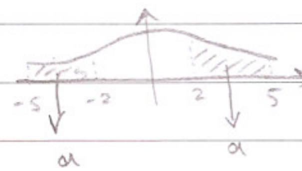
Date: .... Subject: ....

$$\mathbb{P}(2 < |\bar{X}_1 - \bar{X}_2| < 5) = \mathbb{P}(2 < \bar{X}_1 - \bar{X}_2 < 5) + \mathbb{P}(-5 < \bar{X}_1 - \bar{X}_2 <$$

$$= \mathbb{P}\left(\frac{-5-0}{\sqrt{3.4}} < \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < \frac{-2}{\sqrt{3.4}}\right) + \mathbb{P}\left(\frac{2}{\sqrt{3.4}} < \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right.$$

$$= 2\,\mathbb{P}(1.09 < Z < 2.72) = 2(0.9967 - 0.8621)$$

$$= 0.2692.$$



مثال - فرض کنید در دو جامعه میانگین مصرف روزانه پروتئین به ترتیب 125 و 100 گرم است. اگر مقادیر مصرف روزانه پروتئین در دو جامعه دارای توزیع نرمال با انحراف معیار 15 گرم باشد احتمال اینکه نمونه‌های تصادفی و مستقل 25 تایی از هر جامعه دارای تفاوت بین میانگین‌های نمونه کمتر از 12 گرم باشد را بیابید.

$$\bar{X}_1 - \bar{X}_2 \sim N\left(125-100,\ \frac{15^2}{25}+\frac{15^2}{25}\right) = N(25, 18)$$

$$\mathbb{P}(|\bar{X}_1 - \bar{X}_2| < 12) = \mathbb{P}(-12 < \bar{X}_1 - \bar{X}_2 < 12)$$

$$= \mathbb{P}\left(\frac{-12-25}{3\sqrt{2}} < \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < \frac{12-25}{3\sqrt{2}}\right)$$

$$= \mathbb{P}(-8.72 < Z < -3.06) = \mathbb{P}(Z < -3.06) - \mathbb{P}(Z \leq -8.$$

$$= 0.0011 - 0$$

Date:................................ Subject:..

سؤال - برای مقایسه نمرات امتحان آمار دو کلاس، از هر کلاس یک نمونه تصادفی انتخاب کرده‌ایم.

فرض کنید از کلاس اول یک نمونه 25 تایی انتخاب و فاصله اطمینان 95 درصدی

(12.94 , 17.06) برای میانگین نمرات این کلاس بدست آمده است. از کلاس

دوم یک نمونه 16 تایی انتخاب و یک فاصله اطمینان 90 درصدی (10.25 , 13.75)

برای میانگین نمرات آن بدست آمده است. با فرض نرمال بودن نمرات دو کلاس و

مساوی بودن واریانس‌ها، یک فاصله اطمینان 90 درصدی برای اختلاف میانگین

نمرات دو کلاس بدست آورید. چه نتیجه‌ای می‌گیرید.

کلاس اول : $n_1 = 25$ و $1-\alpha = 0/95$

کلاس دوم : $n_2 = 16$ و $1-\alpha = 0/90$

$$\mu_1 \in \left(\bar{x}_1 - t_{1-\alpha/2}(n_1-1)\frac{S_1}{\sqrt{n_1}} , \bar{x}_1 + t_{1-\alpha/2}(n_2-1)\frac{S_2}{\sqrt{n_2}} \right)$$ می‌دانیم

از جدول می‌دانیم $t_{0/975}(24) = 2.06$

$$\Rightarrow \begin{cases} \bar{x}_1 - 2.66 \dfrac{S_1}{5} = 12.94 \\ \bar{x}_1 + 2.66 \dfrac{S_1}{5} = 17.06 \end{cases} \Rightarrow \begin{cases} \bar{x}_1 = 15 \\ S_1 = 5 \end{cases}$$

AZIN

Date:.............................. Subject:..............................

$$t_{0.95}(15) = 1.75$$

$$\begin{cases} \bar{x}_2 - 1.75 \dfrac{s_2}{4} = 10.25 \\ \bar{x}_2 + 1.75 \dfrac{s_2}{4} = 13.75 \end{cases} \Rightarrow \begin{cases} \bar{x}_2 = 12 \\ s_2 = 4 \end{cases}$$

$$\Rightarrow s_p^2 = \frac{24 \times 5^2 + 15 \times 4^2}{39} = 21.538$$

$$\Rightarrow s_p = 4.641$$

$$t_{1-\alpha/2}(n_1+n_2-2) = t_{0.95}(39) = 1.68$$

$$\mu_1 - \mu_2 \in \left(\bar{x}_1 - \bar{x}_2 - t_{1-\alpha/2}(n_1+n_2-2)\, s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \, , \, \bar{x}_1 - \bar{x}_2 + t_{1-\alpha/2}(n_1+n_2-2)\, s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \right)$$

$$\Rightarrow \mu_1 - \mu_2 \in (0.504 \, , \, 5.496)$$

پس با اطمینان ۹۰ درصد میانگین کلاس اول از کلاس دوم بیشتر است.

AZIN

Date:.................................. Subject:..

مثال- در یک اتوبان در هر شبانه روز بطور متوسط 3 تصادف روی می دهد.

الف- اگر بدانیم در یک شبانه روز حداقل دو تصادف روی داده است، احتمال اینکه در همان شبانه روز حداکثر 3 تصادف روی دهد را بیابید.

ب - احتمال اینکه زمان بین دو تصادف متوالی حداقل 8 ساعت باشد را بیابید.

الف-

X = تعداد تصادف ها در یک شبانه روز $\qquad X \sim P(3)$

الف) $$\mathbb{P}(X \le 3 \mid X \ge 2) = \frac{\mathbb{P}(2 \le X \le 3)}{\mathbb{P}(X \ge 2)}$$

$$= \frac{f_X(2) + f_X(3)}{1 - (f_X(0) + f_X(1))} = \frac{0.4481}{0.8009} = 0.5595$$

ب- Y = زمان بین دو تصادف متوالی در یک شبانه روز

$$X \sim P(\mu) \Rightarrow Y \sim E(1/\mu)$$

$$\Rightarrow Y \sim E(1/3)$$

$$\mathbb{P}(Y \ge 1/3) = \int_{1/3}^{\infty} 3e^{-3y}\, dy = -e^{-3y}\Big|_{1/3}^{\infty} = e^{-1} = 0.3679$$

$1/3$ شبانه روز = 8 ساعت

Date: .... Subject: ....

مثال - فرض کنید بهره‌ی هوشی دانشجویی که تصادفاً از یک دانشگاه انتخاب شده است

دارای توزیع نرمال با میانگین ۱۱۰ و انحراف معیار 20 است. فاصله‌ای به

مرکز میانگین را چنان تعیین کنید که ۵۰ ٪ بهره هوشی دانشجویان این دانشگاه را

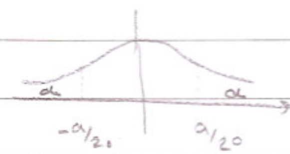
شامل شود. بهره هوشی کلی دانشجو $= X$

$$X \sim N(110 , 20^2)$$

$$\text{فاصله مورد نظر} = (110 - a , 110 + a) \qquad (a>0)$$

$$P(110 - a < X < 110 + a) = 0/5$$

$$= P\left(\frac{-a}{20} < \frac{X-\mu}{\sigma} < \frac{a}{20}\right) = \Phi(a/20) - \Phi(-a/20)$$



$$\Phi\left(-\frac{a}{20}\right) = 1 - \Phi(a/20)$$

$$\Rightarrow 0/5 = 2\Phi(a/20) - 1 \Rightarrow \Phi(a/20) = 0/75$$

$$\Rightarrow \frac{a}{20} = 0/675 \Rightarrow a = 13.5$$

$$\Rightarrow \text{فاصله مورد نظر} = (96.5 , 123.5) .$$

AZIN