Subject: ۱
Date

روش های آماری

امیرحسین کریم زادگان

* مثال: $P(|z| > ۰/۷۸) = P(z > ۰/۷۸) + P(z < -۰/۷۸)$

$= P(z < -۰/۷۸) + P(z < -۰/۷۸)$

$= ۲P(z < -۰/۷۸) = ۰/۴۳۵۴$

* مثال: (فشار خون)

$P(X > ۱۳/۵) \overset{\text{تبدیل = نرمال}}{=} P\left(\frac{X - ۱۱/۵}{\sqrt{۲}} > \frac{۱۳/۵ - ۱۱/۵}{\sqrt{۲}}\right)$

$= P(z > ۱/۴۱) = P(z < -۱/۴۱)$

* مثال: $P(۱۱ < x < ۱۲) = P\left(\frac{۱۱ - ۱۱/۵}{\sqrt{۲}} < \frac{x - ۱۱/۵}{\sqrt{۲}} < \frac{۱۲ - ۱۱/۵}{\sqrt{۲}}\right)$

$= P\left(\frac{-۱}{۲\sqrt{۲}} < \frac{x - ۱۱/۵}{\sqrt{۲}} < \frac{۱}{۲\sqrt{۲}}\right)$

$= P\left(\frac{x - ۱۱/۵}{\sqrt{۲}} < \frac{۱}{۲\sqrt{۲}}\right) - P\left(\frac{x - ۱۱/۵}{\sqrt{۲}} < \frac{-۱}{۲\sqrt{۲}}\right)$

* مثال: الف

$P(X > ۱۶) = P\left(\frac{X - ۱۴/۲۳}{۳} > \frac{۱۶ - ۱۴/۲۳}{۳}\right)$

$= P\left(z > \frac{۱/۷۷}{۳}\right) = P(z > ۰/۵۹) = P(z < -۰/۵۹) = ۰/۲۷۷۶$

ب دو نفر زیر ۱۲ شوند

$P(x < ۱۲) = P\left(\frac{x - ۱۴/۲۳}{۳} < \frac{۱۲ - ۱۴/۲۳}{۳}\right) = P\left(z < \frac{-۲/۲۳}{۳}\right)$

$= P(z < -۰/۷۴) = ۰/۲۲۲۷ \rightarrow (P(x < ۱۲))^۲ = (۰/۲۲۲۷)^۲$

PAPCO

Subject:
Date:

در حالت معمول (وضعیت نرمال)

• مثال: اگر رفتار نمرات یک مدرس دارای میانگین ۱۳٫۸۷ و با انحراف معیار ۱٫۷ را باشد احتمال آن را حساب کنید که متوسط نمره یک کلاس ۲۵ نفره از این شخص:

الف) معدلی برابر با ۱۴٫۲ شود

ب) بیش از ۱۴٫۱۵ شود

ج) حد اکثر ۰٫۰۵ با میانگین همیشگی کلاس های وی اختلاف داشته باشد

الف)
$$P(\bar{X} = 14.2) = 0$$

ب)
$$P(\bar{X} > 14.15) = P\Big(\underbrace{\frac{\bar{X} - 13.87}{\frac{1.7}{\sqrt{25}}}}_{Z} > \frac{14.15 - 13.87}{\frac{1.7}{\sqrt{25}}}\Big)$$

$$= P(Z > 0.8) = P(Z < -0.8) = 0.2119$$

از هر ۱۰۰ کلاس ۲۵ نفره انتظار داریم ۲۱ کلاس معدلی بیش از ۱۴٫۱۵ را داشته باشند.

ج)
$$P(|\bar{x} - 13.87| < 0.05) = P\left(\frac{|\bar{x} - 13.87|}{\frac{1.7}{\sqrt{25}}} < \frac{0.05}{\frac{1.7}{\sqrt{25}}}\right)$$

$$= P(|Z| < 0.15) = P(-0.15 < Z < 0.15)$$

$$= P(Z < 0.15) - P(-0.15 > Z)$$

$$= 1 - P(Z > 0.15) - P(-0.15 > Z)$$

$$= 1 - P(Z < -0.15) - P(Z < -0.15)$$

$$= 1 - 2P(Z < -0.15)$$

$$= 0.1192$$

امیرحسین کریم‌زادگان

Subject: ۲

Date

۱۳ اسفند

- هر چقدر اعضای نمونه بیشتر شود ، میانگین نمونه به میانگین واقعی میل می کند

۱۵ اسفند

- پارامتر : هر مقداری از جامعه که مجهول و غیر قابل محاسبه باشد

- (Mu) μ : میانگین جامعه $$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

- (sigma 2) σ^2: واریانس جامعه $$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X - \mu)^2$$

- (Median) μ_D: میانه جامعه $$\mu_D = X_{(\frac{N}{2})}$$

- ρ_S : چولگی $$\rho_S = \frac{\sum_{i=1}^{N} (X_i - \mu)^3}{N \sigma^3}$$

- ρ_K : کشیدگی $$\rho_K = \frac{\sum_{i=1}^{N} (X_i - \mu)^4}{N \sigma^2}$$

- Q_r (r_{th} quartile) $$Q_r = X_{(1.. r)}$$

- (P_i) π : نسبت در جامعه $$\pi = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{\#(\text{کل افراد جامعه با ویژگی } c)}{N}$$

that : $$Y_i = \begin{cases} 1 & x_i \in C \\ 0 & x_i \notin C \end{cases}$$

- زمان کاربرد پارامترها:

- نسبت در جامعه : ① تعداد زیاد نمونه ② برقرار بودن قضیه حد مرکزی

- واریانس : ① توزیع نرمال

Subject:
Date

- میانگین :

حالت اول : مشاهدات غیرنرمال ، تعداد نمونه زیاد و واریانس معلوم

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

حالت دوم : مشاهدات غیرنرمال ، تعداد نمونه زیاد و واریانس نامعلوم

$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \longrightarrow N(0,1)$$

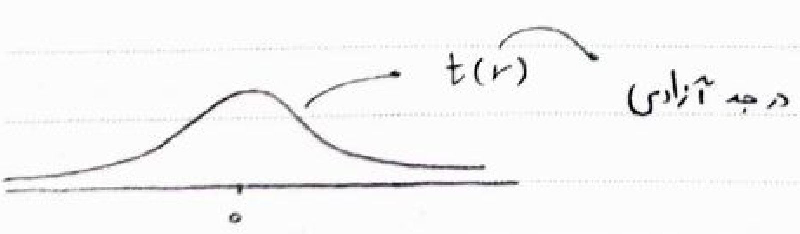
حالت سوم : مشاهدات نرمال و واریانس معلوم

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

حالت چهارم : مشاهدات نرمال و واریانس نامعلوم

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

• توزیع t - استودنت :



$$t(r) \xrightarrow[(r>30)]{r\to\infty} N(0,1)$$

PAPCO

$\boxed{۱۵ \text{ اسفند}}$

• نسبت π :
تعداد نمونه زیاد

$$\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n} \quad , \quad Y_i = \begin{cases} 1 & x_i \in C \\ 0 & x_i \notin C \end{cases}$$

$$\frac{\hat{p} - \pi}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \longrightarrow N(0,1)$$

$$Y_i \sim Ber(\pi) \qquad Var(Y_i) = \pi(1-\pi)$$

• واریانس (σ^2) :
مشاهدات نرمال

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

* χ^2 : (خی ، کی ، کای)

$\boxed{۲۰ \text{ اسفند}}$

• تابع مولد گشتاور :

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_{x \in S_X} e^{tx} P(X=x) \\ \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \end{cases}$$

$$M'_X(t) = E(X e^{tX}) \Big|_{t=0} = E(X) = \begin{cases} \sum_{x \in S_X} x P(X=x) \\ \int_{x \in S_X} x f_X(x)\, dx \end{cases}$$

Subject:
Date

• نتیجه: $M_x^{(n)}(t)\Big|_{t=0} = E(x^n e^{tx})\Big|_{t=0} = E(x^n)$

۱۹ فروردین

برآوردگر	پارامتر
میانگین نمونه $(\bar{X})$	میانگین جامعه (μ)
واریانس (S^2)	واریانس (σ^2) ~
نسبت در نمونه $(\hat{p})$	نسبت در جامعه (π)

• میانگین، ... نمونه توزیع دارد چون از یک مورد به مورد دیگری می‌تواند متفاوت باشد

• میانگین، ... جامعه توزیع ندارد چون ثابت است (یک عدد است)

• وقتی به درصد اطمینان اشاره نشد، آن را ٪ ۹۵ در نظر بگیریم.

۲۱ فروردین

• اگر در سوالی هم واریانس جامعه داده شود هم واریانس نمونه، حتما از واریانس جامعه استفاده می‌کنیم برای حل مسائل

• فاصله اطمینان برای نسبت جامعه:
وقتی پارامتر "نسبت" مجهول است یعنی واریانس جامعه نامعلوم است
↓
یعنی تنها الزام برای امکان ایجاد فاصله اطمینان برای نسبت در جامعه این است که تعداد نمونه زیاد باشد $(n > 30)$
↓
دلیل صفحه بعد

واریانس نمونه در این حالت: $S^2 = \dfrac{\sum x_i^2 - n\bar{X}^2}{n-1} = \dfrac{(n\hat{p}) - n(\hat{p})^2}{n-1}$

PAPCO

تعداد زیاد $\dfrac{n\hat{p} - n(\hat{p})^2}{n} = \hat{p} - \hat{p}^2 = \hat{p}(1-\hat{p})$

۴ | ۲۱ فروردین
Year. Month. Day.

امیرحسین کریم‌زادگان
Subject. ۴۰۱۲۱۸۳۳

- دلیل تساوی صفحه قبل: باید به مفهوم نسبت بپردازیم. در نمونه/جامعه، وقتی به بررسی یک ویژگی می‌پردازیم، اعضا یا این ویژگی را دارند یا ندارند. اگر داشتند به آنها ۱، و اگر نداشتند ۰ می‌دهیم. یعنی هر x_i یا ۰ است یا ۱.

- $\sum x_i^2$: در اینجا داریم x_iهایی که ۱ هستند را ابتدا به توان دو می‌رسانیم سپس آنها را با هم جمع می‌زنیم، این عملا مثل این است که تعداد x_iهایی که ۱ هستند را بیابیم. طبق فرمول $\bar{x} = \hat{p} = \frac{\sum(x_i)}{n}$

که در واقع در صورت کسر داریم تعداد ۰ و ۱ها را جمع می‌زنیم (حاصل صورت همان تعداد ۱ها است)، می‌توانیم تعداد ۱ها را اینگونه بیابیم: $n\hat{P} = \sum(x_i) =$ تعداد ۱ها

پس $\sum x_i^2 = n\hat{P}$

- $\bar{X}$: در مفهوم نسبت، $\hat{P}$ در واقع همان میانگین است

پس $\bar{X} = \hat{P}$

۲۶ فروردین

- میزان پراکندگی می‌تواند { واریانس / انحراف معیار } باشد

- با استفاده از اینکه واحد میزان پراکندگی توان یک دارد یا دو می‌توانیم به واریانس یا انحراف معیار بودن پی ببریم

*مثال: ۲ تن ← انحراف معیار / ۲ تن^2 ← واریانس

- اگر عددی که می‌خواهیم در جدول توزیع موجود نباشد، نزدیک‌ترین عدد به آن را در نظر می‌گیریم.

Year. Month. Day. Subject.

• ادعا مشخص می کند که آزمون فرضیه یک طرفه است یا دو طرفه.

$\boxed{\text{۲ اردیبهشت}}$

• اگر واریانس را نتوان مشخص کرد، بیشترین مقدار داده شده را به عنوان واریانس در نظر می گیریم

$\boxed{\text{۴ اردیبهشت}}$

• $\widehat{\mu_1 - \mu_2} = \hat{\mu}_1 - \hat{\mu}_2 = \bar{X} - \bar{Y}$

• قضیه حد مرکزی برای دو جامعه : $\dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \to N(0,1)$

*اثبات : $\mu^* = \mu_1 - \mu_2$, $\hat{\mu}^* = \bar{X} - \bar{Y} = \bar{Z}$

$$Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y})$$

$$= \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$$

$$\frac{\bar{Z} - \hat{\mu}^*}{\sqrt{Var(\bar{Z})}} = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \longrightarrow N(0,1)$$

• آنچه که از جامعه معلوم است ارجحیت دارد به برآورد.

*مثال : اگر واریانس جامعه (سیگما σ^2) داده شده باشد، با خود آن کار می کنیم، نه با واریانس نمونه (برآورد واریانس S^2)

iman NOTEBOOK

Year. Month. Day.

امیرحسین کرمی ارکان

Subject. ۴۰۱۲۱۸۳۴۳

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right) \qquad \bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n}, \frac{\sigma_2^2}{m}\right)$$

$$\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right) \Rightarrow \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{n}}} \sim N(0,1)$$

$\boxed{\text{۹ اردیبهشت}}$

- واریانس ها نامعلوم و برابر ← برابری واریانس ها در شرایطی رخ می دهد که در جامعه مشابه باشند (از لحاظ پراکندگی)

$$X_1, \dots, X_n \sim N(\mu_1, \sigma^2)$$

$$\frac{(n-1)S_1^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma^2}{n}\right)$$

$$Y_1, \dots, Y_m \sim N(\mu_2, \sigma^2)$$

$$\frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\bar{Y} \sim N\left(\mu_2, \frac{\sigma^2}{m}\right)$$

$$\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma^2}{n}+\frac{\sigma^2}{m}\right)$$

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim N(0,1)$$

$$\frac{(n-1)S_1^2}{\sigma^2}+\frac{(m-1)S_2^2}{\sigma^2} = \frac{(n-1)S_1^2+(m-1)S_2^2}{\sigma^2} \sim \chi^2(m+n-2)$$

(برای هر ترکیب خطی قابل انجام؟)

Year. Month. Day. Subject.

- we know if $Z \sim N(0,1)$, $R \sim \chi^2_{(r)}$

then :

$$\frac{Z}{\sqrt{\frac{R}{r}}} \sim t(r)$$

$$\frac{\dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \left(\frac{1}{n} + \frac{1}{m}\right)}}}{\sqrt{\underbrace{\dfrac{(n-1)S_1^2 + (m-1)S_2^2}{\sigma^2 (m+n-2)}}_{S_p^2}}} = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S_p^2 \left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t(m+n-2)$$

* مثال : حالت ۱ (تعداد نمونه زیاد ، واریانس معلوم) برای $\mu_1 - 2\mu_2$:

$X_1, \dots, X_n$, $E(X_i) = \mu_1$, $Var(X_i) = \sigma_1^2$

$Y_1, \dots, Y_n$, $E(Y_i) = \mu_2$, $Var(Y_i) = \sigma_2^2$

we know $E(\bar{X}) = \mu_1$, $E(\bar{Y}) = \mu_2$

$Var(\bar{X}) = \frac{\sigma_1^2}{n}$, $Var(\bar{Y}) = \frac{\sigma_2^2}{m}$

$E(\bar{X} - 2\bar{Y}) = \mu_1 - 2\mu_2$, so $\bar{X} - 2\bar{Y}$ is an estimator for $\mu_1 - 2\mu_2$.

Based on CLT , we have :

$$\frac{\bar{X} - 2\bar{Y} - (\mu_1 - 2\mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + 4\frac{\sigma_2^2}{m}}} \longrightarrow N(0,1) \quad \text{as } n, m \to \infty$$

*مثال : شهرداری مدعی است افراد ساکن در مناطق بالای شهر مصرف ۱٫۵ برابر افراد ساکن مناطق متوسط شهر دارند.

به منظور بررسی این موضوع میزان مصرف ۷۰ خانوار در مناطق بالای شهر متوسط لیتر مصرفی هر شخص ۸۳ لیتر را نشان می داد. این درحالی است که بررسی ها مربوط به ۵۸ خانوار ساکن در مناطق متوسط شهری ۵۸ لیتر را نشان می دهد. اگر واریانس حاصل از این نمونه ها به ترتیب ۱۲ و ۱۰ لیتر به توان دو باشد فاصله اطمینان برای $2\mu_1 - 3\mu_2$ بدست بیاورید

$$\frac{a\bar{x}+b\bar{y}-(a\mu_1+b\mu_2)}{\sqrt{a^2\frac{\sigma_1^2}{n}+b^2\frac{\sigma_2^2}{n}}} \longrightarrow N(0,1)$$

فاصله اطمینان :

$$(a\bar{X}+b\bar{Y}) \pm Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{a^2\sigma_1^2}{n}+\frac{b^2\sigma_2^2}{m}}\right)$$

$$(2\bar{x}-3\bar{y}) \pm Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{4\sigma_1^2}{n}+\frac{9\sigma_2^2}{m}}\right)$$

$$= (2\times 83 - 3\times 58) \pm Z_{\frac{\alpha}{2}}\left(\sqrt{\frac{4\times 144}{}}\right.$$

۱۶ اردیبهشت

* اثبات قضیه حد مرکزی ← با استفاده از mgf (moment generating function)

• برآورد در داده های زوجی :

هرگاه n نمونه داشته باشیم و اطلاعات حاصل از آنها (مشاهدات) مربوط به دو زمان مختلف به منظور بررسی تاثیر یک روش طراحی شده باشد، عملا از هر نمونه ۲ مشاهده داریم، یعنی این داده ها مستقل نیستند (عدم استقلال مربوط به نمرات [قبل و بعد] یک فرد است)

ادامه صفحه بعد...

Year. Month. Day. Subject.

• این داده ها غالبا حاصل از تکرار یک شاخص اندازه گیری است که به داده های قبل و بعد نیز معروف هستند

شماره نمونه	1	2	3	...	n
مشاهده قبل	X_1	X_2	X_3	...	X_n
مشاهده بعد	Y_1	Y_2	Y_3	...	Y_n
$D_i = Y_i - X_i$	D_1	D_2	D_3	...	D_n

• هدف از جمع آوری این نوع داده ها بررسی اثر بخشی انجام یک روش، استراتژی، روند درمان و غیره می باشد.

عملا به دنبال بررسی $\mu_D = \mu_A - \mu_B$ A: after / B: before

(اینکه کدام را منهای کدام کنیم وابسته به نوع سوال است.)

• جایی که μ_B میانگین قبل از مطالعه و μ_A میانگین بعد از مطالعه موضوع است

• معیار برای قابل بررسی بودن مساله { تعداد نمونه زیاد (حالت ۲) / مشاهدات نرمال (حالت ۴) ← در موارد خاص *

۱- تعداد نمونه زیاد: $\bar{D} \pm Z_{\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$

۲- بررسی فرآیند طبیعی بدون مداخله: $\bar{D} \pm t_{\frac{\alpha}{2}}(n-1) \frac{S_D}{\sqrt{n}}$

* مثال در صفحه بعد (بعدا وارد جزوه شود ...)

* حل سوال ۲ بخش ب میان ترم (بعدا وارد جزوه شود

امیرحسین کریم‌زادگان
Subject ۴۰۱۲۱۸۳۳

* مثال : میزان رشد ۱۲ تخم کدی تغییر ژن داده شده طی ۳ هفته به شکل زیر می باشد :

نمونه	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
میزان رشد	۱۲,۲	۱۳,۴	۱۲,۸	۱۲,۹	۱۳,۵	۱۲,۷	۱۳	۱۳,۲	۱۲,۵	۱۱,۹	۱۲,۴	۱۳,۲

فاصله اطمینان برای میانگین جامعه را بدست آورید :

$\bar{D} = 12,64 \qquad t_{0,025}(11) = 2,201$

$$S = \sqrt{\frac{\sum x_i^2 - n\bar{X}^2}{n-1}} = \sqrt{\frac{1923,69 - 12 \times (12,64)^2}{11}}$$

$$= \sqrt{0,5868} = 0,77$$

$$\Rightarrow 12,64 \pm 2,201 \times \frac{0,77}{\sqrt{11}} = \boxed{(12,13 \,,\, 13,15)}$$

* حل سوال ۲ـ الف میان ترم :

$$\frac{\bar{X}_{1-7} + \bar{X}_{8-12} + \bar{X}_{12-15}}{3} \sim N(10 \,,\, \tfrac{1}{9}$$

$* \; Var(\bar{X}_{1-7}) = \frac{16\, var}{7}$

$$Var(\bar{X}_{1-7} + \bar{X}_{8-12} + \bar{X}_{12-15})$$

$$= Var(\bar{X}_{1-7}) + Var(\bar{X}_{8-12}) + Var(\bar{X}_{12-15}) + 2\,cov(\bar{X}_{8-12} \,,\, \bar{X}_{12-15})$$

$$= \frac{16}{7} + \frac{16}{5} + \frac{16}{4} + 2\left(\frac{16}{20}\right)$$: جواب

$$cov(\bar{X}_{8-12} \,,\, \bar{X}_{12-15}) = \frac{1}{5 \cdot 4} cov(\bar{X}_8 + \bar{X}_9 + \bar{X}_{10} + \bar{X}_{11} + \bar{X}_{12} \,,\, \bar{X}_{12} + \bar{X}_{13} + \bar{X}_{14} + \bar{X}_{15}$$

$$= \frac{1}{20} cov(X_{12} \,,\, X_{12}) = \frac{1}{20} Var(X_{12}) = \frac{1}{20} \cdot \frac{16}{1}$$

$$= \frac{16}{20}$$

۱۸ اردیبهشت

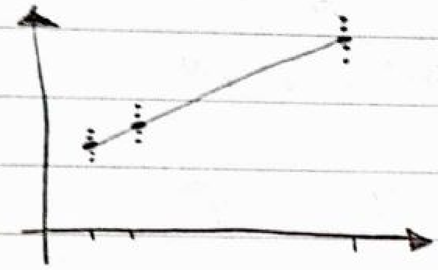
Year. Month. Day. (مطالعه رفتار) Subject.

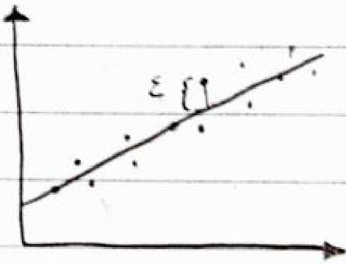
• آزمون های زوجی (جامعه های غیرمستقل) ⇐ بیشتر برای یک گروه درستی یک موضوع است

• آزمون با جامعه های مستقل ⇐ بیشتر برای مقایسه دو چیز با هم است

• رگرسیون خطی ساده : (X_i , Y_i) $\quad i = 1, \dots, n$ $\qquad |P_{x,y}| \in (0.85 , 1]$

متغیر پاسخ، وابسته — متغیر کنترلی، توضیحی — ضریب همبستگی





$$Y = a + b X + \underbrace{\varepsilon}_{\text{Error term}} \qquad E(Y) = a + bX$$

$$Y = E(Y) + \varepsilon$$

خطا دارای یک توزیع با میانگین برابر با 0 است $\qquad \varepsilon \sim (0 , \sigma^2)$

• در مورد خطا هیچ اطلاعاتی به جز این ۳ نداریم:

۱) میانگین خطا صفر است

۲) دارای واریانس ثابت می باشد (منظور پراکندگی است، واحد نیست. واحدها می توانند متفاوت باشند ولی پراکندگی در همه واحدها تقریبا مشابه است)

۳) خطا ها مستقل از یکدیگرند

در ریاضی

$E(Y) = a + bX$

عرض از مبدا — شیب

در آمار

• a : متوسط مقدار Y به ازای صفر بودن X می باشد (میانگین) (?)

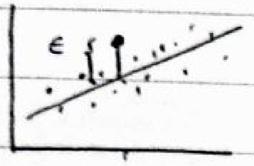
• b : متوسط تغییر در Y به ازای یک واحد تغییر در X

Year. Month. Day.

امیرحسین کریم زادگان

Subject. ۴۰۱۲۱۸۳۳

- خطا در مدل رگرسیونی : $\epsilon_i = y_i - (a + bx_i)$



برای یافتن ϵ ، نقطه را به خط عمود نمی کنیم !

- گاهی اوقات نمودار پراکندگی رفتار خطی ندارد و رفتارش از درجه بیش از یک است :

* یک ایده : Segmentation regression (تقسیم شکل به چند بخش و نمایش هر بخش با یک رابطه خطی)

- برای اینکه بفهمیم یک نمودار از درجه چندم است : باید ببینیم چند بار تغییر جهت در آن رخ داده ← (درجه نمودار) = 1 + (تعداد تغییر جهت ٤)

* برای یافتن LSE (Least Square error) چون به دنبال یافتن min هستیم باید معادلات را نسبت به هر کدام از متغیر ٤ مشتق بگیریم

- مثبت یا منفی شدن عرض از مبدا در این موثر که بفهمیم کدام مورد مستقل است و کدام وابسته

۱ خرداد

Year. Month. Day. Subject.

- If $\varepsilon_i \sim N(0, \sigma^2)$

then $\hat{b} \sim N\left(b, \frac{\sigma^2}{SSX}\right) \xRightarrow{\text{برای استاندارد شدن}} \frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} \sim N(0,1)$

& $\hat{a} \sim N\left(a, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SSX}\right)\right) \xRightarrow[\text{برای استاندارد شدن}]{} \frac{\hat{a}-a}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{SSX}\right)}} \sim N(0,1)$

- $c, d > 0$

$P(\hat{b} - c < b < \hat{b} + d) = 1-\alpha$

$\hookrightarrow P(-\hat{b} - d < -b < -\hat{b} + c) = P(-d < \hat{b} - b < c)$

$\hookrightarrow = P\left(\frac{-d}{\sqrt{\frac{\sigma^2}{SSX}}} < \frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} < \frac{c}{\sqrt{\frac{\sigma^2}{SSX}}}\right)$

$\hookrightarrow = P\left(-d^* < \frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} < c^*\right)$

$\Rightarrow c^* = d^* = Z_{\frac{\alpha}{2}} \rightsquigarrow c = d = Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{SSX}}$

۶ خرداد

- $E(y) = a + bx$

- b: میزان تغییر در مقدار y به ازای یک واحد تغییر در متغیر x

- $\begin{cases} H_0: & b = b_0 \\ H_1: & b \neq b_0 \end{cases}$

$$\begin{cases} H_0: & b = b_0 \\ H_1: & b < b_0 \end{cases}$$

Year. Month. Day. ۹

امیرحسین کریم‌زادگان

Subject. ۴۰۱۲۱۸۳۳

- ایجاد ناحیه رد برای b :

$$c < b_0 < d$$

$$R = \{ \hat{b} < c \text{ یا } \hat{b} > d \}$$

$$\alpha = P(R \mid H_0) = P(\hat{b} < c \text{ یا } \hat{b} > d \mid b = \hat{b}\}$$

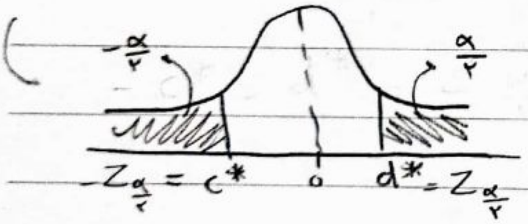
$$= P_{b=b_0}(\hat{b} - b < c - b \text{ یا } \hat{b} - b > d - b\}$$

$$= P_{b=b_0}\left(\frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} < \frac{c-b}{\sqrt{\frac{\sigma^2}{SSX}}} \text{ یا } \frac{\hat{b}-b}{\sqrt{\frac{\sigma^2}{SSX}}} > \frac{d-b}{\sqrt{\frac{\sigma^2}{SSX}}} \right)$$

در این مرحله $b = b_0$ اعمال می‌شود و دیگر لازم نیست بنویسیمش در مراحل بعد

$$= P(Z < c^* \text{ یا } Z > d^*)$$

$$= P(Z < c^*) + P(Z > d^*)$$



$$\longrightarrow \quad -c^* = d^* = Z_{\frac{\alpha}{2}}$$

$$\Rightarrow R = \left\{ \frac{\hat{b}-b_0}{\sqrt{\frac{\sigma^2}{SSX}}} < -Z_{\frac{\alpha}{2}} \text{ or } \frac{\hat{b}-b_0}{\sqrt{\frac{\sigma^2}{SSX}}} > Z_{\frac{\alpha}{2}} \right\}$$

$$= \left\{ \left| \frac{\hat{b}-b_0}{\sqrt{\frac{\sigma^2}{SSX}}} \right| > Z_{\frac{\alpha}{2}} \right\}$$

• ایجاد ناحیه رد یک طرفه برای b :

$$\begin{cases} H_0 : & b = b_0 \\ H_1 : & b < b_0 \end{cases}$$

$$c < b_0$$

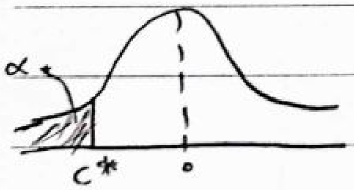
$$R = \{ \hat{b} < c \}$$

$$\alpha = P(R \mid H_0) = P(\hat{b} < c \mid b = b_0)$$

$$= P_{b=b_0}\left(\hat{b} - b < c - b \right)$$

$$= P_{b=b_0}\left(\frac{\hat{b} - b}{\sqrt{\frac{\sigma^2}{SSx}}} < \frac{c - b}{\sqrt{\frac{\sigma^2}{SSx}}} \right)$$

$$= P(z < c^*)$$



$\longrightarrow \quad c^* = -z_\alpha$

$$\Rightarrow \quad R = \left\{ \frac{\hat{b} - b_0}{\sqrt{\frac{\sigma^2}{SSx}}} < -z_\alpha \right\}$$

امیرحسین کریم‌زادگان
Subject. ۴۰۱۲۱۸۳۳

• یافتن ناحیه رد (دو طرفه) برای a :

$$\begin{cases} H_0 : a = a_0 \\ H_1 : a \neq a_0 \end{cases}$$

$$c < a_0 < d$$

$$R = \{ \hat{a} < c \quad \text{or} \quad \hat{a} > d \}$$

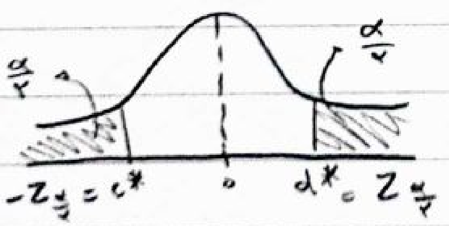
$$\alpha = P(R|H_0) = P(\hat{a} < c \text{ or } \hat{a} > d \mid a = a_0)$$

$$= P_{a=a_0}(\hat{a} - a < c - a \quad \text{or} \quad \hat{a} - a > d - a)$$

$$= P_{a=a_0}\left(\frac{\hat{a}-a}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{SSx}\right)}} < \overset{\ominus}{\frac{c-a}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{SSx}\right)}}} \quad \text{or} \quad \frac{\hat{a}-a}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{SSx}\right)}} > \overset{\oplus}{\frac{d-a}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{SSx}\right)}}} \right)$$

$$= P(Z < c^* \quad \text{or} \quad Z > d^*)$$

$$= P(Z < \overset{\ominus}{c^*}) + P(Z > \overset{\oplus}{d^*})$$



$$\longrightarrow \quad -c^* = d^* = Z_{\frac{\alpha}{2}}$$

$$R = \left\{ \frac{\hat{a}-a}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{SSx}\right)}} < -Z_{\frac{\alpha}{2}} \quad \text{or} \quad \frac{\hat{a}-a}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{SSx}\right)}} \right\}$$

$$R = \left\{ \left| \frac{\hat{a}-a}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{SSx}\right)}} \right| > Z_{\frac{\alpha}{2}} \right\}$$

۸ خرداد

Year. Month. Day. Subject.

- $E(Y) = a + bx$

  $y = a + bx + ۴$

- $\varepsilon \sim N(0, \sigma^2)$ ① ذات مشاهدات ② ابزار های تشخیص

- (x_i, y_i) $i = 1, \dots, n$

  $y \stackrel{?}{=} a + bx + \varepsilon_i$

① رسم نمودار پراکندگی

② بررسی توزیع جملد خطا ⟵ تنها به بررسی نرمال بودن می پردازیم

③ بررسی توزیع جملد خطا

④ برآورد a, b

⑤ آزمون فرضیه درباره a, b

* جملات خطا معلوم نیستند ⟵ برای دیدن آنها یک مفهوم جدید تعریف می‌کنیم ⟵ باقی مانده

$e_i = Y_i - \hat{Y}_i$ & $\hat{Y}_i = \hat{a} + \hat{b} x_i$

وقتی گفته می‌شود توزیع $\hat{y}_i$ را بیابید، چون می دانیم $\hat{y}_i = \hat{a} + \hat{b}x_i$ و $\hat{a}$ و $\hat{b}$ توزیع نرمال دارند، پس می فهمیم که $\hat{y}_i$ توزیع نرمال دارد، پس کافی است برویم سراغ یافتن میانگین و واریانس ($E(x)$ و $\sigma^2(x)$)

Subject:
Date

• یافتن توزیع $\hat{y}_i$:

$$\hat{y}_i = \hat{a} + \hat{b}x_i$$

$$= (\bar{y} - \hat{b}\bar{x}) + \hat{b}x_i$$

$$= \bar{y} + \hat{b}(x_i - \bar{x})$$

$$= \left(\sum_{j=1}^{n} \frac{1}{n} y_j\right) + \left(\sum_{j=1}^{n} \frac{(x_j - \bar{x})}{SSX} \cdot y_j\right)(x_i - \bar{x})$$

$$= \sum_{j=1}^{n} \left(\frac{1}{n} + \frac{(x_j - \bar{x})}{SSX}(x_i - \bar{x})\right) y_j$$

که لازم است چند این ادامه یابد تا $E(\hat{y}_i)$ و $Var(\hat{y}_i)$ را بیابیم.

• جواب نهایی :

$$\hat{y}_i \sim N\left(a + bx_i \,,\, \sigma^2\left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{SSX}\right)\right)$$

• $E(\hat{y}_i) = E(\hat{a} + \hat{b}x_i) = E(\hat{a}) + E(\hat{b}x_i)$

$$= E(\hat{a}) + E(\hat{b})x_i$$

$$= a + bx_i$$

?

• $Var(\hat{y}_i) = Var(\hat{a} + \hat{b}x_i)$

*داریم :

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \Rightarrow Var(\hat{a}) = \left(\frac{1}{n} + \frac{\bar{x}^2}{SSX}\right)\sigma^2$$

$$\hat{y}_i = \bar{y} + \hat{b}(x_i - \bar{x}) \Rightarrow Var(\hat{y}_i) = \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{SSX}\right)\sigma^2$$

Subject:
Date

• برای یافتن واریانس $\hat{Y}_i$ روند صفحه قبل را ادامه دهیم:

$$\hat{Y}_i = \sum_{j=1}^{n} \left(\frac{1}{n} + \frac{(x_j - \bar{x})}{SSx} (x_i - \bar{x}) \right) Y_i$$

مقدار $\left(\frac{1}{n} + \frac{(x_j-\bar{x})}{SSx} (x_i - \bar{x}) \right)$ در Y_i ضرب شده پس در واریانس این ضریب توان ۲ می گیرد.

$$Var(\hat{Y}_i) = \sum_{j=1} \left(\frac{1}{n} + \frac{(x_j - \bar{x})}{SSx} (x_i - \bar{x}) \right)^2 Y_i$$

؟

۱۳ خرداد

• $\varepsilon \sim N(0, \sigma^2)$

$$e_i = Y_i - \hat{Y}_i \sim N\left(0, \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{SSx} \right)\right)$$

$$E(e_i) = E(Y_i) - E(\hat{Y}_i) = (a + bx_i) - (a + bx_i) = 0$$

• $Y_i \sim N(a + bx_i, \sigma^2)$

$$\bar{Y} \sim N(?, ?)$$

$$\bar{Y} = \frac{\sum Y_i}{n}$$

مدل رگرسیونی: $\left\{ \begin{array}{l} E(Y_i) = a + bx_i \\ Y_i = a + bx_i + \varepsilon_i \end{array} \right\}$

• $E(\bar{Y}) = \frac{1}{n} E\left(\sum Y_i \right) = \frac{1}{n} \sum E(Y_i) = \frac{1}{n} \sum E(a + bx_i)$

$= a + b\bar{x}$

PAPCO

CamScanner

۱۳ خرداد

Year. Month. Day. Subject.

• $Var(\bar{y}) = Var\left(\frac{1}{n}\sum y_i\right) = \frac{1}{n^2} Var\left(\sum y_i\right)$

$$= \frac{1}{n^2}\sum Var(y_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

• Mean Square error (MSE):

$$MSE = \frac{\sum_{i=1}^{n} e_i^2}{n-2} \sim \frac{\sigma^2}{n-2}\chi^2_{(n-2)}$$

$$MSE = \frac{SSE}{n-2}$$

• نکته : $E(\chi^2_{(r)}) = r$

$$\Rightarrow E(MSE) = E\left(\frac{SSE}{2}\right) = \frac{\sigma^2}{n-2}E\left(\chi^2_{(n-2)}\right) = \sigma^2$$

• اگر فرض $\varepsilon_i \sim N(0, \sigma^2)$ برقرار باشد، آنگاه MSE یک برآوردگر مناسب برای پارامتر مجهول σ^2 می باشد

* درصورتی که σ^2 معلوم باشد نیازی به برآورد کردن نداریم. از خود مقدار σ^2 استفاده می کنیم.

• واریانس خطا همیشه در واقعیت مجهول است (زیرا خود خطا مجهول است)

Year. Month. Day. Subject.

- $\hat{b} \sim N\left(b, \frac{\sigma^2}{SSX}\right)$

$\hat{a} \sim N\left(a, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SSX}\right)\right)$

$MSE \sim \frac{\sigma^2}{n-2} \chi^2_{(n-2)} \iff \frac{(n-2)\,MSE}{\sigma^2} \sim \chi^2_{(n-2)}$

- If $Z \sim N(0,1)$, $U \sim \chi^2_{(r)}$ are ind,

then $\frac{Z}{\sqrt{\frac{U}{r}}} \sim t_{(r)}$

- $\frac{\hat{b} - b}{\sqrt{\frac{\sigma^2}{SSX}}} \sim N(0,1)$

$$\Rightarrow \frac{\frac{\hat{b} - b}{\sqrt{\frac{\sigma^2}{SSX}}}}{\sqrt{\frac{\frac{(n-2)\,MSE}{\sigma^2}}{(n-2)}}} = \frac{\hat{b} - b}{\sqrt{\frac{MSE}{SSX}}} \sim t_{(n-2)}$$

- $\frac{\hat{a} - a}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SSX}\right)}} \sim N(0,1)$

$$\Rightarrow \frac{\hat{a} - a}{\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{SSX}\right)}} \sim t_{(n-2)}$$