

اینجانب علیرضا درویشی متعهد می شوم در پاسخ به نفع به تعقالات این آزمون مانند طبعیات حضوری امتحان عمل کنم.

1. به ازای هر بار بیشتر و زیاده کمتر در طول دوران بارداری ~~0.00663~~ به طر میانگین 0.663 در هر

وزن نوزاد بیشتری شود ~~0.00663~~

$\frac{\beta}{SE(\beta)} = 5$ پس ضریب از نظر آماری معنی دار است. چون t لغبت از 2 بیشتر

است و می توان با سطح اطمینان 95٪ فرض صفر بودن β را رد کرد.

انتظار می رود که ضریب مثبت باشد و با مراجعات بیشتر به پزشک، کودک سالم تر باشد و وزن بیشتری داشته باشد. که این اتفاق افتاده است.

2. $H_0: \beta = 0.005$

$H_1: \beta \neq 0.005$

$t\text{-score} = \frac{0.00663 - 0.005}{0.00132} = 1.235 \rightarrow P\text{-value} = 0.216 > 0.05$

پس فرض صفر رد نمی شود. یعنی دلایل آماری کافی برای اینکه نشان بدهیم فرض $\beta = 0.005$ برقرار نیست

وجود ندارد و فرض $\beta = 0.005$ پذیرفته می شود

3.

اگر اثر علی وجود داشته باشد، 4 فرض برقرار هستند:

1- هم تابع واقعی خطی است: $\ln b w g h t_i = \beta_0 + \beta_1 n p v i s + u_i$

2- داده گیری به شکل تصادفی است و داده ها مستقل هستند

3- ~~داده ها~~ متغیر توصیف کننده ثابت نیست

4 - $E(U|npv3) = 0$ یعنی متغیرهای دیگری که توضیح دهنده از $npv3$ مستقل هستند

اگر اثر علی برقرار باشد، می فهمیم که رانندگی علییت بین متغیر توضیح دهنده و Y برقرار است و می توانیم به مادران توصیه کنیم که برای دانش فرزندشان با وزن بیشتر، بیشتر به دختر مراجعه کنند.
ولی هم بستگی می تواند اتفاقی باشد. می تواند به دلیل یک تغییر معذوب باشد. می تواند دلیل علیت معلوم باشد یعنی دلیل مراجعه به بیشتر مادران، وزن بیشتر نوزادشان باشد. با وجود همبستگی، نمی توان توضیحی به مادران کرد.

4 شرط گفته شده برقرار هستند یا خیر:

- شرط اول می تواند درست باشد. حداقل با تقریب خیلی، درست است.

- شرط دوم لگش به داده گیری دارد و احتمالاً درست است.

- شرط سوم به دلیل متغیر مزبور وزن دارایی $npv3$ برقرار است.

- شرط 4: متغیرهای معذوب زیادی می تواند وجود داشته باشد که بررسی نشده. برای مثال درآمد خانواده.

با تعداد مراجعات رایج دارد. از طرفی مادر با درآمد بیشتر، ویتامین های بیشتری مصرف می کند.

یا حتی متغیری مثل خرج روی مکه های غذایی هم متغیر می باشد.

4. ضریب متغیر $npv3$ منفی است و از نظر آماری معنی دار است و طبق پیش بینی درست است.

مادران به صرف بسیار بیشتر، کودک نا سالم تری دارند.

ضریب متغیر $media$ معنی دار نیست. می توان حدس زد ضریب باید مثبت باشد. چون مادران تحصیل کرده تر،

مراقبت های پزشکی بیشتری انجام می دهند.

ضریب $mage$ معنی دار است نیست. در من انتظار دام ^{متنی} باشد. معنی مادران کم سن تر سالم تر هستند و بچه های سالم تری هم دارند.

ضریب $male$ معنی دار است و مثبت است. این یعنی ^{متنی} نوزادان پسر به طور میانگین age وزن بیشتری دارند که اثری مطابق پیشبینی است.

5. بخش اول توضیح دهنده $npv5$ در ستون 1، به دلیل هم بستگی $npv5$ با متغیرهای اضافه شده است. برای مثال فرض کنید $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ و واقعی باشد اما از وگرسیون مدل متغیر استفاده کنیم $y = \beta_0 + \beta_1 x$ حال می توان به راحتی دید:

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \delta$$

که δ از همبستگی x_1 با x_2 ^{متنی} به دست می آید:

$$x_2 = \alpha + \delta x_1$$

و $\hat{\beta}_1$ تورش خواهد داشت. در این مثال خاص:

- مادرانی که می بایستی باشند احتمالاً به طور بیش از حدی برای کودکان خود قابل نیستند و به دکتر هم زیاد مراجعه نمی کنند. یعنی همبستگی تعداد بچه ها و تعداد مراجعه وجود دارد.
- مادران با سن بیشتر، شاید دلیل ندرای بیشتر، بیشتر به دکتر مراجعه می کنند.
- مادران با تحصیلات بیشتر به دلیل مشغله ای بیشتر، کمتر به دکتر مراجعه می کنند.

6. با اضافه شدن متغیر age توضیح دهنده $npv5$ مدل بیشتر شده است. در این صورت باید داده

افزایش در $npv5$ $1.25 + 0.0267 \cdot npv5 \cdot x_2$ درصد وزن نوزاد بیشتری شود و باید واحد افزایش $mage$

$0.0329 \cdot mage \cdot x_2 - 2.08$ درصد وزن نوزاد بیشتری شود. در این صورت با تغییر از 11 به 12 تعداد ویرس

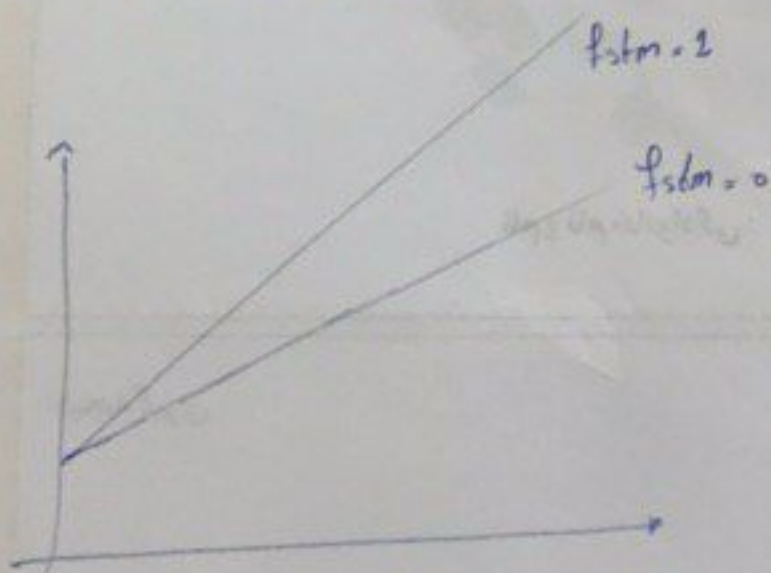
0.63 درصد وزن نوزاد افزایش می یابد.

7.

تفسیر ضریب β مادرانی که در ماه اول به پزشک مراجعه کرده اند: خود یا اولین نوزادی - دنیای آلوده
 2.99 درصد وزن کمتری دارند. علامت ضریب تفاوت با پیشینه است و انتظار می رود مادران
 با $\beta = 1$ ، نوزادان سالم تری داشته باشند. یکی از دلایل این عدم تطابق پیشینه با واقعیت
 می تواند این باشد که مادرانی که ماه اول مراجعه کرده اند، مشکل داشته اند یا فرزند مشکل داری داشته اند که در ماه اول
 مراجعه کرده اند. پس این ~~مشکل~~ نوزادان وزن کمتری دارند.

8.

فرض شده است که مادرانی که در ماه اول به دکتر مراجعه کرده اند، به ازای هر بار مراجعه بیشتر، دکتر،
 تاثير متفاوتی روی وزن نوزاد می پذیرند. اگر ضریب بدست آمده مثبت باشد، فرضیهی بخش قبل را بهتر می توان
 پذیرفت. ضریب 0.00891 یعنی مادران مراجعه کننده به دکتر در ماه اول، ~~بیشتر~~
 نوزادی با وزن 0.891 درصد وزن بیشتر به ازای هر مراجعه بیشتر، دکتر خواهند داشت. این یعنی
 مشکلی که باعث شده مادر در ماه اول مراجعه کند، ~~بیشتر~~ با مراجعه به دکتر برطرف می شود.



مادران زیر توضیح بیشتری می دهد.

9. R^2 به معنی کسری از تغییرات (Inbought) است که مدل توضیح می دهد. هر چه R^2 به 1 نزدیک تر باشد، مدل توضیح دهنده بهتری دارد و به خط ~~بیش~~ پخش شدنی شده نزدیک است. در واقع:

$$R^2 = \frac{SSE}{SST}, \quad SST = \sum (y_i - \bar{y})^2, \quad SSE = \sum (\hat{y}_i - \bar{y})^2$$

\bar{R}^2 معیاری است که R^2 را با تعداد پارامترها تنظیم می کند. در واقع هر چه پارامترهای بیشتری برای مدل انتخاب کنیم، توضیح دهنده مدل بیشتری شود اما امکان دارد مدل overfit کند. \bar{R}^2 معیاری است که R^2 را با تعداد پارامترها تنظیم می کند تا از overfit در زمان انتخاب مدل جلوگیری کند.

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}$$

که k تعداد متغیرهای توضیح دهنده است.

کوچک بودن R^2 ، \bar{R}^2 به معنی توضیح دهنده بسیار کم هر ستون است. هر چند هم چنان می تواند رابطی منطقی برقرار باشد و دلیل کوچک بودن R^2 ، بزرگ بودن واریانس u و کم بودن متغیرهای توضیح دهنده در مدل باشد.

ستون 5 و \bar{R}^2 بیشتری دارد بین این مدل ها، بهترین است.

۱۰- آماره F را برای این مدل می توان بدست آورد.

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{k, n-k-1}$$

$$\left\{ \begin{array}{l} H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1: \text{حداقل یکی از } \beta\text{ها مخالف صفر باشد} \end{array} \right\} \quad \text{آزمون فرض:}$$

$$F = \frac{\frac{0.0189}{5}}{\frac{1-0.0189}{1764-5-1}}$$

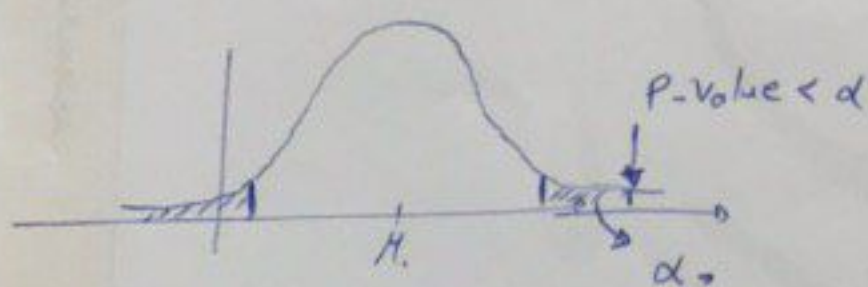
+ توزیع این آماره توزیع F با درجات آزادی k در $n-k-1$ است.

۱۱-

P -value در یک تست یعنی احتمال خطای نوع ۱ در صورت رد فرض صفر. به این معنی که اگر فرض صفر درست

باشد و فرض صفر را رد کنیم، با احتمال P -value اشتباه کرده ایم. حل اگر P از ۵٪ کمتر باشد،

با سطح اطمینان ۰.۹۵ می توان فرض صفر را رد کرد. در شکل زیر می توان آزمون فرض $H_0: \mu = \mu_0$ را دید.



یعنی با خطای بسیار کمی می توان فرض صفر را رد کرد. اگر آزمون استفاده شده همان آزمون بغش قبلی

باشد و فرض صفر، $\beta_1 = \beta_2 = \dots = \beta_k = 0$ باشد، معنی این P -value این است که حداقل یکی از β ها

خالصه معنی دار است.

12. تاثیر $npvis$ در وزن: $0.5\% = 1.2 - 2 \times 0.0306 \times 11.6$

هر واحد اضافه ویزیت، 0.5 درصد وزن بیشتر برای ماندن با $p_{stmon} = 0$ در پی خواهد داشت.

$$\ln\left(\frac{mean + std}{mean}\right) = 16\%$$

این نتیجه ^{خیلی خوب نیست} ~~چون~~ ^{چون}

یعنی با حدود 30 بار مراجعی بیشتر، می توان در حد یک std وزن بیشتری برای هم پیدا کرد. نتیجه برای

$p_{stmon} = 1$ ، بهتر است و از نظر اقتصادی به صرفه تر هم هست. (30 بار مراجعی بیشتر یعنی حدود 5 بار ~~بیشتر~~ برابر

std مراجعات)

بسیار

$$p_{stmon} = 1 : 0.5 + 0.9 = 1.4$$

با حدود 10 بار مراجعی بیشتر، وزن به تعداد به اندازه ای یک std زیاد می شود که بسیار خوب است.

و حدود 3 برابر تعداد مراجعات است.

std