

اینجانب علیرضا درویشی مقوم می‌شود در پاسخ به لایحه اعتراض این آراء مندرج در جلسات خصوصی امتحان عملی.

1. به ازای هر بار بیشتر و بیشتر دقت در طول دوران بارداری ~~0.00663~~ به ~~0.00663~~ 0.00663 و وزن نوزاد بیشتری شود.

$$\frac{\beta}{SE(\beta)} = 5$$

این ضریب از نظر آماری معنی دار است چون t نسبت از 2 بیشتر

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

انتظار می‌رود که ضریب مثبت باشد و با مراجعات بیشتر به پزشک، کودک سالم‌تر باشد و وزن بیشتری داشته باشد که این اتفاق افتاده است.

$$H_0: \beta = 0.005$$

$$H_1: \beta \neq 0.005$$

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

پس فرض صفر رد نمی‌شود. یعنی دلایل آماری کافی برای اینکه نشان بدهیم فرض $\beta = 0.005$ برقرار نیست وجود ندارد و فرض $\beta = 0.005$ پذیرفته می‌شود.

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

$$\ln b_{wght} = \beta_0 + \beta_1 \ln pvis + u$$

1- هم تابع واقعی خطی است.

2- داده‌گیری به شکل تصادفی است و داده‌ها مستقل هستند.

3- ~~داده‌ها~~ متغیر توصیف کننده ثابت نیست.

①

4 - $E(U1 \cdot p_{13}) = 0$ یعنی متغیرهای دیگری که توابع تصحیحی در آن هستند از p_{13} مستقل هستند.

اگر اثر علی برقرار باشد، می فهمیم که θ رابطه‌ی علیت بین متغیر توابع تصحیحی و γ برقرار است و می توانیم به مادران توصیه کنیم که برای دانش فرزندی با وزن بیشتر، بیشتر و کمتر مراجعه کنند.
ولی هم بستگی می تواند اتفاقی باشد. می تواند دلیل رد تاثیر محدود باشد می تواند دلیل علیت معلوم باشد یعنی دلیل مراجعتی بیشتر مادران، وزن بیشتر نورادشان باشد. با وجود همبستگی می توان توصیه ای به مادران کرد.

4 شرط کوفه شده برقرار هستند یا خیر:

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

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

- شرط سوم به دلیل غیر همبستگی واریانس p_{13} برقرار است.

- شرط 4 متغیرهای محدود ریاضی می تواند وجود داشته باشد که بررسی نشده. برای مثال درآمد خانوار.

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

یاضی متغیری مثل خرج روی مکمل های غذای هم متغیر می باشد.

4- ضریب متغیر 50% منفی است و از نظر آماری معنی دار است و طبق پیش بینی درست است.

مادران کم مصرف سیگار بیشتر، کودک ناسالم 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.

(2)

ضریب ρ_{age} معنی دار آمار نیست. و من ρ انتظار دارم ρ_{age} معنی دار باشد. معنی مادران کم سن تر، سالم تر هستند و بعد های سالم تر می دارند.

ضریب ρ_{age} معنی دار است و مثبت است. این یعنی ρ_{age} نوزادان بزرگتر - طور میانگین ρ_{age} وزن بیشتری دارند که انرژی موافق پیش بینی است.

5. بعد از توضیح دهنگی ρ_{age} در ستون 1، به دلیل هم بستگی ρ_{age} 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:

$$E(\hat{\beta}_1) = \beta_0 + \beta_2 x$$

که x از همبستگی x_2 با x_1 ρ_{age} بدست می آید.

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

- مادرانی که سیگار می کشند احتمالاً به طور بیش از حدی برای کودک خود قائل نیستند و به دلتر هم زیاد مراجعه نمی کنند. یعنی همبستگی تعداد سیگار و تعداد مراجعه وجود دارد.

- مادران با سن بیشتر، شاید به دلیل نگرانی بیشتر، بیشتر به دکتر مراجعه می کنند.

- مادران با تحصیلات بیشتر به دلیل مستعدی بیشتر، کمتر به دکتر مراجعه می کنند.

6. با انجام تست همبستگی معکوس شده، توضیح دهنگی مدل بیشتر شده است. در این صورت باید داده

افزایش در ρ_{age} $1.25 + 0.0267 \cdot \rho_{age} x_2$ درصد وزن نوزاد بیشتری شود و باید واحد افزایش ρ_{age}

$\rho_{age} x_2$ $0.0329 - 2.08$ درصد وزن نوزاد بیشتری شود. در این صورت با تغییر از 11 به 12 تعداد ویریت

0.63 درصد وزن نوزاد افزایش می یابد.

7.

تفسیر ضریب: مادرانی که در ماه اول به پرسنک مراجعه کرده اند، خود می دانند که نوزادی که در دنیا آمده است

2.99 درصد وزن کمتری دارند. علامت ضریب متفاوت با پیش بینی است و انتظار می رود مادران

با $\beta_{slmon} = 1$ ، نوزادان سالم تری داشته باشند. یک دلیل این عدم تطابق پیش بینی 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 ~~نوزادان~~ نوزادان وزن کمتری دارند.

8.

فرض شده است که مادرانی که در ماه اول به داکتر مراجعه کرده اند، به ازای هر بار مراجعه ای بیشتر داکتر،

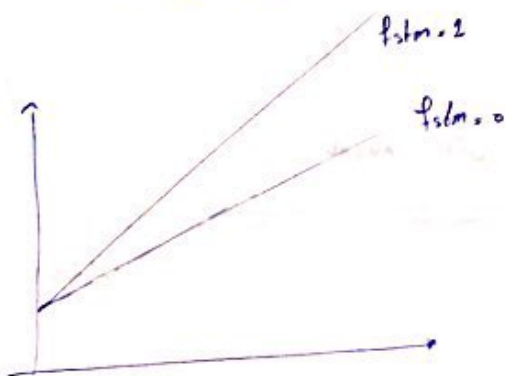
تاثير متفاوتی روی وزن نوزاد می پذیرند. اگر ضریب بدست آمده مثبت باشد، فرضیه ی بخش قبل را بهتری توان

پذیرفت. ضریب 0.00891 یعنی مادران مراجعه شده به داکتر در ماه اول، ~~نوزادان~~

نوزادی با وزن 0.891 درصد وزن بیشتر به ازای هر مراجعه ای بیشتر - داکتر خواهند داشت. این یعنی

مشکلی که باعث شده مادر در ماه اول مراجعه کند، ~~نوزاد~~ با مراجعه به داکتر برطرف می شود.

بنابراین زیر توضیح بهتری می دهد.



R^2 به معنی کسری از تغییرات (variance) است که مدل توضیح می دهد. هر چه R^2 به 1 نزدیکتر باشد،

مدل توضیح دهنده بهتری دارد و به خط ~~خط~~ بیش از حد فتنه 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

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 ، به معنی توضیح دهنده بسیار کم هر ستون است. هر چه هم چنان می تواند رابطه ی

برقرار باشد و دلیل کوچک بودن R^2 ، بزرگ بودن واریانس ϵ و کم بودن متغیرهای توضیح دهنده در مدل باشد.

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

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

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

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

$$F = \frac{0.0189}{5} \div \frac{1-0.0189}{1764.5-1}$$

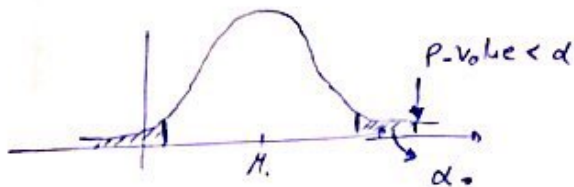
توزیع این آماره توزیع F با درجات آزادی k و $n-k-1$ است.

۱۱-

$P\text{-value}$ در یک تست یعنی احتمال خطای نوع ۲ در صورت رد فرض صفر به این معنی که اگر فرض صفر درست

باشد و فرض صفر را رد کنیم، با احتمال $P\text{-value}$ اشتباه کرده ایم. حل ما P از ۵٪ کمتر باشد،

با سطح اطمینان ۹۵٪ می توان فرض صفر را رد کرد. در شکل زیر می توان آزمون فرض $H_0: H_1$ 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 $P\text{-value}$ این است که حداقل یکی از β ها نامصفره معنی دار است.

12. تاثیر η_{pvis} در وزن: $0.5\% = 1.2 - 2 \times 0.0306 \times 11.6$

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

این نتیجه ^{خیلی خوب نیست} ~~چون~~ $\ln\left(\frac{mon + std}{mon}\right) = 16\%$

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

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

$0.5 + 0.9 = 1.4\%$: $p_{slmon} = 1$

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

و حدود 3 برابر تعداد مراجعات است.