پیش‌بینی غیبت کارکنان با استفاده از دو روش رگرسیونی کمترین مجذورهای متداول و توبیت

Predicting Employee Absenteeism Using Tobit Regression and OLS Models

دکتر حجت‌الله فراهانی
روان‌شناس

دکتر حمیدرضا عریضی
استادیار دانشگاه اصفهان

دکتر محمدکاظم سلیمی‌زاده
استادیار دانشگاه علامه طباطبایی

Hojjattollah Farahani, PhD
Psychologist

Hamid R. Oreizy, PhD
Isfahan University

Mohammad K. Salimizadeh, PhD
Allame Tabatabaee University

چکیده

یافته‌های پـژوهشی حاصل از روش رگرسیون کمترین مجذورهای متداول (OLS) دربـاره غیبت کارکنان، به‌ویژه به‌دلیل همبستـگی متغیرهای پیش‌بینی‌کننده مـورد انتقاد قـرار گرفتـه است. در این پژوهش نتایـج پیش‌بینی غیبت کارکنان بر پایه چهار متغیر دلبستگی شغلی، غیبت در مقایسه بادیگران، تنیدگی رویدادهای زندگی و تعداد فرزندان، با دو روش رگرسیونی کمترین مجذورهای متداول (OLS) و تحلیل توبیت (Tobit) مقایسه شد. ۱۹۷ نفر از کارکنان مرد کارخانه ذوب آهن و فولاد مبارکه اصفهان (با میانگین سنی ۲۸ و انحراف استاندارد ۱۰ سال) به مقیاسهای فهـرست تنیدگی رویدادهای زندگی تویتس (۱۹۸۱)، غیبت در مقایسه بـا دیگران (بابا، ۱۹۹۰)، دلبستگی شغلی (لودال و کجنر، ۱۹۶۵) پاسخ دادند. تحلیل توبیت، غیبت شغلی را نسبتاً دقیق‌تر و با ۱۳ درصد تبیین واریانس بیشتر (R^2=۴۲/۱) پیش‌بینی کرد.

واژه‌های کلیدی : رگـرسیون کمترین مجذورهای متـداول، توبیت، غیبت کارکنان.

Abstract

The results of predicting employee absenteeism based an Ordinary Least Square (OLS) has been criticized. The present research compared the results from predicting absenteeism based on four varibales of job involvement, absenteeism compared to others, stress from life events, and number of children with the OLS and Tobit regression models. A sample of 197 male employees of factories in Esfahan city mean age : 28, SD : 10 responded to the Life Events Stress Inventory (Thoits, 1981), The Absenteeism Compared to Others Inventory (Baba, 1990), and the Job Involvement Inventory (Lodahl & Kejner, 1965). Tobit model analysis predicted employee absenteeism more precisely and explained a furhter 13% of the variance (R^2 = 42.1) than did OLS

Key words: Ordinary Least Squares regression, Tobit, absenteeism.

دریافت : ۸۶/۸/۲۳ received : 13 Nov 2007

پذیرش : ۸۷/۹/۱۷ accepted : 7 Dec 2008

Contact information : icphaf@yahoo.com

مقدمه

موضوع غیبت شغلی همواره مورد توجه و علاقه روان‌شناسان صنعتی و سازمانی بوده است. غیبت شغلی عبارت است از حضور کمتر افراد در کار، در حالی که ساعتهای حضور موظف افراد طبق برنامه از قبل تعیین شده است (جانز، ۲۰۰۰ نقل از موسوی، ۱۳۸۴).

غیبت می‌تواند ارادی و ناشی از نارضایتی شغلی فرد باشد و یا از دلایل غیرارادی، غیرقابل اجتناب و موجه مانند بیماری ناشی شود (احمدیان، ۱۳۷۶).

تعیین میزان غیبت مجاز برای سنجش غیبت شغلی به‌عنوان ملاک موفقیت شغلی چندان مناسب نیست (جانز، ۱۹۹۴ و ۱۹۸۷ نقل از موسوی، ۱۳۸۴). چون کارکنان غالباً دلایل غیبت خود را به مشکلات پزشکی کوچک اسناد می‌دهند و صحت این اسناد مورد تردید است و از سوی دیگر غیبت به دلیل بیماری با متغیرهای انگیزشی همبستگی دارد، به‌همین دلیل نوع غیبت تا حد زیادی بازتاب هنجارهای اجتماعی است (جانز، ۱۹۹۴ نقل از موسوی، ۱۳۸۴).

کارکنان گاهی نقش افراد بیمار را ایفا می‌کنند تا از این طریق غیبت کنند (گیگ، ۱۹۸۳). از سوی دیگر ممکن است افرادی که باید واقعاً به دلیل بیماری غیبت کنند بر سرکار خود حاضر شوند (گیبسون، ۱۹۹۶). این موضوع بیش حضوری[1] نامیده می‌شود و میزان واقعی غیبت به‌دلیل بیماری را تغییر می‌دهد. برای بررسی این متغیر شیوه‌های آماری مختلفی وجود دارد که با توجه به هدفهای پژوهش تعیین می‌شوند، اما همواره یکی از مسائل مهم در پژوهشهای علمی شیوه‌ای است که از آن برای پاسخگویی به مسئله پژوهشی استفاده می‌شود. این شیوه نوع نتایج حاصل را تحت تأثیر قرار می‌دهد (کوهن، ۱۹۷۰). هنگامی که به مفروضه‌های مدل آماری مورد استفاده توجه شود، آماره‌های حاصل معمولاً پایا و کارا خواهند بود (گرین، ۱۹۹۳) یا به‌عبارت دیگر، روابط واقعی خود را نشان خواهند داد. وقتی از این مفروضه‌ها عدول می‌شود، مانند هنگامی که از رگرسیون کمترین مجذورهای متداول[2] (OLS) برای داده‌های غیرنرمال استفاده می‌شود برآورد حاصل ممکن است دقیق نباشد چون عدم توجه به این مفروضه‌ها، نادیده گرفتن روابط موجود یا شناسایی یک رابطه کاذب و تصنعی را در پی دارد.

یکی از حیطه‌های مهم در پژوهشهای منابع انسانی که از نظر روش شناختی حساسیت دارد، پژوهش در حوزه غیبت شغلی کارکنان است. عمده‌ترین مشکل پژوهشهای مبتنی بر منابع انسانی این است که میزان کارکنان از توزیع نرمال تبعیت نمی‌کند (اربوس و سیچل، ۱۹۵۴؛ بابا، ۱۹۹۰؛ هامر و لاندائو، ۱۹۸۱؛ لاندائو، ویس و اسمیت، ۱۹۸۴). هامر و لاندائو (۱۹۸۱) برخی از مسائل روش‌شناختی مربوط به پژوهشهای انجام شده در مورد غیبت کارکنان را مورد بحث و بررسی قرار داده و معتقدند از آنجایی که در سنجش غیبت مقدار منفی وجود ندارد و فراوانی بالای غیبت اندک است، توزیع در اکثر مواقع دارای چولگی شدید است و نمی‌توان برای تحلیل آن از روش OLS استفاده کرد.

مادالا (۱۹۸۴) نشان داده که جمله خطا با متغیرهای مستقل همبستگی دارد و در نتیجه برآورد OLS اریب[3] خواهد بود. بنابراین برای پیش‌بینی متغیری مانند غیبت کارکنان استفاده از مدلی توصیه می‌شودکه بتوان از طریق آن برآورد پایا[4] و دقیقی از غیبت به دست آورد. روش Tobit برای این موقعیت توصیه می‌شود زیرا پیش‌بینی مقادیر زیر صفر و اریب جانبی[5] OLS را ندارد و نیز کارآیی[6] آن از OLS برای همان توزیع بیشتر است (لیگ، ۱۹۸۵؛ آناستازسوپایلوس، تروکو و مامرینگ، ۲۰۰۷).

رگرسیون کمترین مجذورهای متداول (OLS) یکی از روشهای رایج در پیش‌بینی یک متغیر وابسته (ملاک) براساس ترکیب خطی از متغیرهای پیش‌بین (مستقل) است. در همه مسائل رگرسیونی اعم از ساده یا چند متغیری از اصل کمترین مجذورها استفاده می‌شود. پیش‌بینی هر متغیر بر مبنای متغیرهای دیگر همیشه با خطا همراه است. بنابراین نه‌تنها داده‌های علمی با این خطا همراهند بلکه در این میان داده‌های علوم رفتاری بیش از داده‌های حاصل

1. presenteeism
2. Ordinary Least Squares regression
3. biased
4. stable
5. asymptotic bias
6. efficiency

از علوم طبیعی در معرض این خطای آماری قرار می‌گیرند. به‌عبارت دیگر، خطاهای پیش‌بینی در تحلیل، بزرگتر و بارزترند یا واریانس خطا بزرگتر است. در حالی که با اصل کمترین مجذورها می‌توان داده‌ها را چنان تحلیل کرد که مجذورهای خطای پیش‌بینی به کمترین حد برسند (کرلینگر[1] و پدهازر[2]، ۲۰۰۱/۱۳۸۴).

رگرسیون چندگانه (OLS) خواص الگوی رگرسیون ساده را دارد. در رگرسیون چندگانه تعداد متغیرهای پیش‌بین معادله از ۱ تا P افزایش می‌یابد و معادله آن را می‌توان به‌صورت زیر در نظر گرفت:

$$\hat{y} = b_1x_1 + b_2x_2 + b_3x_3 + ... + b_kx_k + a + e$$

a، عرض از مبدأ معادله رگرسیون و بخشی از y را هنگامی که مقادیر تمامی متغیرهای پیش‌بین صفر هستند، تشکیل می‌دهد. هرگاه تمام متغیرها x_k (برای $(k=1,...,p)$) مرکزی شده باشند، مقدار آن برابر میانگین y یعنی $\hat{y}$ است. در این معادله $b_1, b_2, ..., b_k$ شیب خط یا ضرایب خط رگرسیون را نشان می‌دهند. ضرایب معادله رگرسیون قابلیت تفسیر دارند، مثلاً اگر تمام متغیرهای پیش‌بین x_2 تا x_k ثابت نگه داشته شوند، هر یک واحد افزایش در x_1, b_2 واحد افزایش در y را در بر می‌گیرد (وی‌آن و اسکاستر، ۱۹۹۸/۱۳۸۱).

در تفسیر نمره b باید از واحد اندازه‌گیری آگاه بود، زیرا تغییر در واحدهای اندازه‌گیری b ها را تغییر می‌دهد. به همین دلیل در معادله رگرسیون β نمایانگر تغییر مورد انتظار در متغیر وابسته‌ای است که به‌صورت نمره استاندارد بیان می‌شود و به‌ازای یک انحراف استاندارد تغییر در میزان متغیر مستقل در شرایطی که سایر متغیرهای مستقل کنترل یا ثابت نگه داشته شوند به وجود می‌آید (سرمد، بازرگان و حجازی، ۱۳۸۰). بر این اساس در مورد یک گروه N نفری، N پیش‌بینی $\hat{y}$ وجود دارد. به‌عبارت دیگر، در نمره ملاک هر فرد از طریق نمره پیش‌بین وی که مشاهده شده و موجود است پیش‌بینی می‌شود. همانگونه که گفته شد این کار با خطا همراه است. بر مبنای اصل کمترین مجذورها، y ها را باید چنان پیش‌بینی کرد که مجموع مجذورهای خطا در پیش‌بینی یعنی $\sum (y_i - \hat{y}_i)^2$ به کمترین برسد. در این فرمول عدد y شماره آزمودنی و y_i نمره مشاهده شده او در متغیر وابسته و $\hat{y}_i$ نمره پیش‌بینی شده او در آن متغیر است. در واقع اصل کمترین مجذورها مبتنی براین است که a و ضرایب b باید به گونه‌ای محاسبه شوند که خطا در پیش‌بینی به کمترین برسد. یا به عبارت دیگر، اصل کمترین مجذورها تحقق یابد (کرلینگر و پدهازر، ۲۰۰۱/۱۳۸۴). از سوی دیگر β برخلاف b تحت تأثیر واحد اندازه‌گیری متغیرها قرار نمی‌گیرد (فراهانی و عریضی،۱۳۸۴). β خاص داده‌های نمونه محاسبه شده است و قابل تعمیم به شرایط و جوامع دیگر نیست. مقدار b به‌رغم تفاوت واریانسها و کوواریانسها در شرایط متفاوت ثبات بیشتری نشان می‌دهد. توصیه اغلب متخصصان این است که هر گاه هدف مقایسه اثر متغیرهای متفاوت در یک جامعه واحد است باید از β و هر گاه هدف مقایسه اثر متغیرهای خاص در جوامع متفاوت است باید از b استفاده شود (فراهانی و عریضی، ۱۳۸۴). می‌توان از طریق فرمول زیر b را به β تبدیل کرد:

$$\beta = \frac{Sx}{Sy} b$$

برای آزمون اینکه آیا پارامترهای برآورد شده در رگرسیون چندگانه به‌طور معناداری با صفر اختلاف دارند یا خیر؟ پژوهشگران از آزمون t یا F استفاده می‌کنند (وی‌آن[3] و اسکاستر[4]، ۱۳۸۱/ ۱۹۹۸). با این حال ممکن است وزنهای β به‌خودی‌خود اطلاعات چندانی در اختیار قرار ندهند. بنابراین ضریب رگرسیون چندگانه محاسبه می‌شود که عبارت از مجذور ضریب همبستگی بین y و بهترین ترکیب خطی متغیرهای پیش‌بین است. این ضرایب با R^2 نمایش داده می‌شوند. R^2 سهم متغیرهای مستقل را در تبیین واریانس متغیر وابسته نشان می‌دهد. به عبارت دیگر درصد واریانس تبیین شده متغیر وابسته توسط متغیرهای مستقل R^2 و از طریق فرمول $\frac{SSreg}{Sst}$ محاسبه می‌شود (فراهانی و عریضی، ۱۳۸۴). معناداری مقدار R^2 از طریق محاسبه نسبت F طبق فرمول زیر مشخص

1. Kerlinger, F. N
2. Pedhazure, E. G.
3. Vieon, R.
4. Schuster, C.

می‌شود:

$$F=\frac{R^2/K}{(1-R^2)/(N-K-1)}$$

(K : تعداد متغیرهای مستقل)

از آنجا که R^2 برآورد ناااریبی از پارامتر (ρ_{xy}) نیست و مقدار آن اریب بستگی به اندازه نسبی P , n دارد، برآورد ناااریبی از آن توسط فرمولی خاص محاسبه و تحت عنوان R^2 تعدیل شده[1] آورده می‌شود.

لازم به توضیح است که برای رسیدن به مدل رگرسیون کمترین مجذورهای متداول (OLS) روشهای متفاوتی به کار می‌روند. معادله رگرسیون با روش ورود[2]، گام به گام[3]، حذف پس‌رونده و یا روش سلسله مراتبی[4] قابل محاسبه و دستیابی است.

هر یک از این روشها اهمیت خاص خود را دارند. برای مثال در روش گام‌به‌گام اهمیت نسبی متغیرهای پیش‌بین تعیین و معضل هم‌خطی چندگانه[5] کنترل می‌شود.

استفاده از رگرسیون چندگانه وابسته به شرایطی است تا تحلیل بتواند برآوردی پایا، دقیق (با کمترین خطا) و کاراً ارائه دهد. این مفروضه‌ها فقط خاص این روش نیستند و در بسیاری از روشهای آماری چند متغیری مانند تحلیل واریانس چند متغیری[6] و تحلیل کوواریانس[7] و ... باید لحاظ شوند. یکی از نکات مهم در استفاده از این روش آماری شناسایی داده‌های دور افتاده[8] (پرت) است. از آنجا که رگرسیون روش معروف به بیشینه‌سازی آماری[9] است، بنابراین در پی «بهترین برازش[10]» برای مسئله است، در نتیجه این روش نسبت به داده‌های دورافتاده بسیار حساس است. این داده‌ها نمره‌های افراطی هستند که ممکن است در متغیرهای ملاک (پیش‌بینی شونده) و پیش‌بین دیده شوند، به‌طور بارزی از y های واقعی فاصله دارند و می‌توان آنها را با استفاده از آزمون ویس‌برگ[11] که معناداری آن با آزمون t انجام می‌شود بررسی کرد. نمره‌های دورافتاده در متغیرهای پیش‌بین نیز می‌توانند در تحلیل رگرسیون OLS مشکل ایجاد کنند. این مسئله را می‌توان با استفاده از آماره معروف به فاصله ماهالانوبیس[12] تعیین کرد. این آماره نشان می‌دهد که هر فرد از سنتروئید[13] متغیرهای پیش‌بین چقدر فاصله دارد. معناداری این آماره را می‌توان از طریق آماره مجذور کای (χ^2) آزمون کرد (فراهانی و عریضی، ۱۳۸۴). گاهی داده‌های دورافتاده از طریق آماره‌های مذکور قابل بررسی نیستند، در این موقع آماره‌ای موسوم به فاصله کوک[14] نشان می‌دهد که اگر در رگرسیون یک مورد خاص حذف و رگرسیون مجدداً انجام شود، ضرایب رگرسیون چه مقدار تغییر خواهد کرد. علاوه بر نقش داده‌های دورافتاده در مدل OLS باید به اهمیت هم‌خطی چندگانه توجه ویژه‌ای داشت. اگر دو متغیر همبستگی بالایی با یکدیگر داشته باشند، احتمال دارد که آنها دقیقاً واریانس مشابهی را در y تبیین کنند، این موضع هم‌خطی چندگانه نامیده می‌شود (فراهانی، ۱۳۸۲).

این امر در رگرسیون چندگانه از طریق آماره تحمل[15] و عامل تورم واریانس[16] (VIF) قابل بررسی است. تحمل یک متغیر پیش‌بین نشان می‌دهد که آن متغیر تا چه حد می‌تواند توسط دیگر متغیرهای پیش‌بین موجود در الگو پیش‌بینی شود. این آماره میزان همپوشی متغیرهای پیش‌بین را نشان می‌دهد. هر چه میزان این ضرایب بالاتر باشد، میزان همپوشی با دیگر متغیرهای پیش‌بین کمتر است. افزون بر آن، این ضریب با ثبات الگو در ارتباط است. مقدار پایین آن، ثبات الگو و دقت در محاسبه را به مخاطره می‌اندازد (هومن، ۱۳۸۰).

همبستگی بالای متغیرهای پیش‌بین با یکدیگر در هم‌خطی چندگانه، خطای استاندارد ضرایب معادله را متورم می‌سازد و در نتیجه از دقت برآورد کاسته می‌شود، این مسئله به عامل تورم واریانس موسوم است که باید در مواقعی که مقدار آن بزرگ (بیشتر از ۲) می‌شود مسئله رابطه بین متغیرهای پیش‌بین بررسی شود (فراهانی و عریضی، ۱۳۸۴).

1. adjusted
2. enter
3. stepwise
4. hierarchical regression
5. multicolinearity
6. Multivariate Analysis of Variance
7. Analysis of Covariance
8. outlier data
9. statistical maximization
10. best fit
11. Weisberg's outlier data
12. Mahalanobis distance
13. centroid
14. Cook's distance
15. tolerance
16. variance inflative factor

روشهای متفاوتی برای حذف هم‌خطی چندگانه وجود دارند. استفاده از ماتریس همبستگی بین متغیرهای پیش‌بین، حذف یا ترکیب متغیرهای دارای همبستگی بالا، رگرسیون کمکی[1]، رگرسیون گام‌به‌گام و تبدیل لگاریتمی از آن جمله‌اند (فراهانی، ۱۳۸۲).

نکته مهم دیگر که در روش رگرسیون OLS باید مورد توجه قرار گیرد، بررسی نمودار مانده‌ها[2] است. مقادیر مانده یا خطا در رگرسیون به تفاوت $\hat{y}, y$ یا به عبارت دیگر مقدار دقت پیش‌بینی معادله رگرسیون اشاره دارد. در صورت غیرنرمال بودن، غیرخطی‌بودن و ناهمگنی پراکنش نمره‌های مانده، بهتر است تحلیل مناسب‌تری را جایگزین رگرسیون خطی (OLS) کرد. غیرنرمال‌بودن نشان می‌دهد که الگوی منحنی، درجه دوم است که در این حالت باید از تحلیل مناسب رگرسیون غیرخطی و لوجستیک استفاده شود. بنابراین برای آنکه رگرسیون OLS بتواند به اصل کمترین مجذورهای خطا دست یابد، باید داده‌ها مورد بررسی قرار گیرند و رگرسیون خطی OLS که به‌طور متداول در علوم رفتاری از آن استفاده می‌شود زمانی برآوردی دقیق و پایا ارائه خواهد داد که مفروضه‌های مطرح شده مورد بررسی قرار گیرند.

مدل دیگری که برای برآورد استفاده می‌شود و در برخی از موارد مناسب‌تر از مدل OLS است، مدل Tobit است که در علوم رفتاری کمتر مورد توجه قرار گرفته است. مدل Tobit (توبین، ۱۹۵۸) به مدلهای رگرسیونی اشاره دارد که در آن دامنه متغیر وابسته در برخی ازموارد سانسور شده* است (وانگ، ۲۰۰۷). این مدل که نام آن از مجموع حرف اول توبین (اولین کسی که این مدل را معرفی کرد) و چهار حرف آخر پروبیت (Probit) تشکیل شده است؛ یک روش رگرسیونی است که برای داده‌هایی به‌کار می‌رود که دارای بخش گسسته و پیوسته‌اند. در روشهای مختلف رگرسیونی یا داده‌های تحلیل گسسته‌اند یا پیوسته، اما در مدل Tobit الگویی از ترکیب هر دو نوع داده وجود دارد. توبیت را می‌توان گسترش روش پروبیت و رویکردی مناسب برای برخورد با داده‌های سانسور شده دانست (جانستون و دیناردو، ۱۹۹۷). سانسورسازی داده‌ها یک محدودیت داده‌ای است که در داده‌های گردآوری شده از سطوح پایین‌تر یا بالاتر از آستانه و یا هر دو حالت ممکن با هم، به وقوع می‌پیوندد (آناستاسو-پایلوس و دیگران، ۲۰۰۷). به‌عبارت دیگر، هنگام استفاده از رگرسیون OLS برای برخی از داده‌ها، مقادیر منفی پیش‌بینی می‌شود که به طور واضح بی‌معنا هستند. برای مثال، تصمیم فرد برای خرید یک اتومبیل در یک دقیقه خاص را در نظر بگیرید و متغیر y را به عنوان شاخص تمایل فرد برای خرید یک اتومبیل تلقی کنید. در این حالت فرد یا اتومبیل را خریداری می‌کند یا از خرید آن صرف‌نظر می‌کند. بنابراین، امکان منفی‌شدن y وجود ندارد، اگر برای این پیش‌بینی از الگوی OLS استفاده شود ممکن است مقادیر منفی به‌دست آیند که بی‌معنا هستند. در مورد غیبت کارکنان نیز وضعیت اینگونه است. بنابراین Tobit یک مدل رگرسیونی مناسب برای داده‌های بریده شده با احتمال بالا است که متغیر فراتر از آن نقطه به طور پیوسته توزیع می‌یابد. به عبارت دیگر ویژگی اصلی مدل توبیت متغیر ملاک با احتمال بالا برای برخی از مقادیر است (در مورد غیبت کارکنان این مقدار صفر است، زیرا اکثر کارکنان غیبت غیرمجاز نمی‌کنند) و توزیع آن غیر از مقدار با احتمال بالا به صورتی پیوسته است. برآوردهای به‌دست آمده با استفاده از این مدل از ثبات، اعتبار بیشتر و اریب کمتر در مقایسه با مدل OLS برخوردارند (بابا، ۱۹۹۰؛ لیگ، ۱۹۸۵؛ مادالا، ۱۹۸۴). در این روش نیز مانند دیگر مدلهای لگاریم خطی[3] از روشهای احتمال بیشینه برای برآورد وزنهای معادله رگرسیون استفاده می‌شود و معادله نهایی نشان‌دهنده توزیع چوله با تابع چگالی نرمال استاندارد است (هامر و لاندائو، ۱۹۸۱). مفروضه‌های غیرخطی معادله ضریب رگرسیون استاندارد β در مدل Tobit وجود ندارد اما نسبت t اطلاعاتی درباره سهم متغیر مستقل مفروض در پیش‌بینی متغیر ملاک فراهم می‌کند.

1. ridge regression
2. residual plot
3. log linear

* منظور داده‌هایی هستند که در توزیع آنها پس از یک مشاهده مشخص (x_i) سایر مشاهده‌ها حاصل نشده است.

از این طریق می‌توان پیش‌بینیهای مدل OLS و Tobit را مقایسه کرد. در مدل Tobit دو جمعیت واجد و فاقد چولگی در نظر گرفته می‌شود. برای اولی، ضریب β محاسبه می‌شود. این امر امکان بررسی دقیق را فراهم می‌آورد، زیرا یک توزیع چوله[1] منجر به ناهمگنی پراکنش[2] می‌شود که به شدت خطای استاندارد را تحت‌الشعاع قرار می‌دهد و به خطای نوع اول یا دوم منتهی می‌شود (هامر و لاندائو، ۱۹۸۱).

هامر و لاندائو (۱۹۸۱) نیز برای اجتناب از کم برآورد کردن[3] رابطه مفروض بین غیبت و پیش‌بینی‌کننده‌های آن، مدل Tobit را پیشنهاد کرده‌اند. از این مدل برای پیش‌بینی غیبت کارکنان در پژوهشهای مختلف استفاده شده است. برای مثال لیگ (۲۰۰۴) برای پیش‌بینی غیبت کارکنان براساس متغیرهای جنس، فشار زیاد[4] و خطرهای شغلی در نمونه ۱۳۰۸نفری از شاغلان با بیش از ۲۰ ساعت کار در ایالت متحده پژوهشی انجام داد و به دلیل ماهیت متغیر غیبت از روش Tobit استفاده کرد. هامر و لاندائو (۱۹۸۱) و بابا (۱۹۹۰) در پژوهشهای خود نشان دادند که مدل Tobit نسبت به مدل رگرسیونی OLS در پیش‌بینی غیبت کارکنان حساس‌تر است.

این پژوهش به مقایسه دو مدل رگرسیونی مطرح شده، در پیش‌بینی طول مدت غیبت کارکنان کارخانه ذوب آهن اصفهان و فولاد مبارکه براساس ۴ متغیر مؤثر در غیبت می‌پردازد. فرض اصلی این پژوهش این است که مدل Tobit در پیش‌بینی غیبت کارکنان نسبت به مدل OLS دقت وکارآیی بیشتری دارد.

روش

گروه نمونه را ۱۹۷ نفر از کارکنان مرد کارخانه ذوب آهن اصفهان و فولاد مبارکه در سال ۱۳۸۴ با میانگین سنی ۲۸ و انحراف استاندارد ۱۰ سال تشکیل دادند. ۸۴ درصد آنها دارای مدرک کارشناسی و کمتر و ۱۶ درصد دارای مدرک کارشناسی ارشد و بالاتر بودند.

فهرست تنیدگی رویدادهای زندگی تویتس (۱۹۸۱) : این فهرست از ۱۰ گویه با طیف سه درجه‌ای (بسیار زیاد=۳ و کم =۱) تشکیل شده است. اعتبار این ابزار از طریق آلفای کرونباخ در این پژوهش برابر با ۰/۸۶ به دست آمد.

غیبت در مقایسه با دیگران (بابا، ۱۹۹۰) : با استفاده از این پرسشنامه ۲۰ سؤالی غیبت در مقایسه با دیگران بررسی شد. این ابزار تأثیر غیبت فرد را در مقایسه با غیبت دیگران در محیط کار می‌سنجد. پاسخهای این پرسشنامه با دامنه ۱-۳ نمره‌گذاری می‌شوند. اعتبار این پرسشنامه از طریق آلفای کرونباخ برابر ۰/۸۳ به دست آمد.

دلبستگی شغلی از طریق مقیاس فرم کوتاه ۶ گزاره‌ای لودال[5] و کجنر[6] (۱۹۶۵) در طیف لیکرت ۵ درجه‌ای از کاملاً موافق (۱) تا کاملاً مخالف (۵) اندازه‌گیری شد. اعتباراین ابزار از طریق آلفای کرونباخ ۰/۸۷ به دست آمد. برای داده‌های حاصل نشده (مفقود) از متوسط نمره‌های هر مقیاس استفاده شد.

غیبت کارکنان نیز براساس کل روزهای غیبت در ۹ ماه آخر کار کارکنان براساس پرونده‌های موجود سنجیده شد. انتخاب دوره ۹ ماهه در سنجش غیبت به‌منظور اطمینان از ثبات در اندازه‌های غیبت بود. برای پیش‌بینی مدت زمان غیبت کارکنان براساس متغیرهای پیش‌بین از دو مدل OLS و Tobit استفاده شد و این دو مدل مورد مقایسه قرار گرفت.

یافته‌ها

جدول ۱ ضریبهای همبستگی بین متغیرهای پیش‌بین مدل را نشان می‌دهد. در جدول ۲، نتایج حاصل از تحلیل OLS برای پیش- بینی طول مدت زمان غیبت منعکس شده است.

بررسی چهار متغیر پیش‌بینی‌کننده مدت زمان غیبت کارکنان با استفاده از معادله رگرسیون معمولی (OLS) نشان داد که بیشترین ضریب β مربوط به غیبت در مقایسه با دیگران است و چهار متغیر این تحلیل در

1. skewed distribution
2. heteroscedasticity
4. underestimation
5. hyperpressure
5. Lodahl, T. M.
6. Kejner, M.

جدول ۱.
ضریبهای همبستگی بین متغیرهای مورد مطالعه

متغیرها	۱	۲	۳	۴
۱- تعداد فرزندان	—			
۲- دلبستگی شغلی	۰/۱۲	—		
۳- غیبت در مقایسه با دیگران	۰/۲۱*	-۰/۲۴*	—	
۴- تنیدگی رویدادهای زندگی	۰/۲۵*	-۰/۱۱	۰/۱۸	—
۵- غیبت شغلی	۰/۲۲*	-۰/۴۲*	-۰/۳۱*	۰/۲۷*

*P<۰/۰۵.

جدول ۲.
خلاصه تحلیل رگرسیون طول مدت زمان غیبت با روش کمترین مجذورهای معمولی (OLS)

متغیر پیش‌بین	β	b	SE	R^2
تعداد فرزندان	۰/۲۹*	-۱/۲۴*	۰/۲۷	۰/۰۸۴
دلبستگی شغلی	-۰/۳۱**	۱/۴۹	۰/۴۱	۰/۱۸۰
غیبت در مقایسه با دیگران	-۰/۲۷*	-۱/۵۳	۰/۶۷	۰/۲۵۲
تنیدگی رویدادهای زندگی	۰/۱۹*	۰/۹۲	۰/۴۱	۰/۲۸۹

*P< ۰/۰۵. **P< ۰/۰۱.

جدول ۳.
خلاصه تحلیل رگرسیون طول مدت زمان غیبت با روش Tobit

متغیر پیش‌بین	β	$\hat{\beta}$	SE	b	R^2
تعداد فرزندان	۰/۲۹*	-۱/۶۲	۰/۲۹	-۱/۲۷	۰/۲۲۱
دلبستگی شغلی	۰/۳۱**	-۲/۲۲	۰/۵۷	-۱/۶۲	۰/۲۸۱
غیبت در مقایسه با دیگران	-۰/۲۷**	-۲/۱۷	۰/۷۷	-۱/۱۱	۰/۳۰۲
تنیدگی رویدادهای زندگی	۰/۱۹*	۱/۳۲	۰/۴۹	۰/۸۴	۰/۴۲۱

توجه : $\hat{\beta}$ وزن ضرایب مدل، Tobit، با فرض چوله بودن توزیع *P< ۰/۰۵. **P< ۰/۰۱.

مجموع ۲۸/۹ درصد واریانس را تبیین کردند.

جدول ۳ نتایج حاصل از تحلیل Tobit برای پیش‌بینی طول مدت زمان غیبت شغلی را نشان می‌دهد.

نتایج بررسی چهار متغیر تعداد فرزندان، دلبستگی شغلی و غیبت در مقایسه با دیگران، تنیدگی رویدادهای زندگی در پیش بینی طول مدت زمان غیبت کارکنان با استفاده از مدل Tobit (جدول ۳) نشان داد که چهار متغیر مورد مطالعه نسبت به مدل OLS واریانس بیشتر و برابر با ۴۲/۱ درصد را تبیین می‌کند. بنابراین، روش Tobit نسبت به OLS واریانس بیشتری برای داده‌های غیبت فراهم می‌آورد.

بحث

نتایج حاصل از این پژوهش نشان دادند که هنگام استفاده از مدل Tobit در مقایسه با OLS، برآورد دقیق‌تری در پیش‌بینی غیبت شغلی کارکنان براساس متغیرهای پیش‌بینی کننده غیبت فراهم آورده می‌شود و نیز واریانس بیشتری از متغیر ملاک براساس متغیرهای پیش‌بین تبیین می‌شود که این با نتایج پرایس و مولر (۱۹۸۶) که به نتایج مشابهی با نتایج این پژوهش دست یافته‌اند، هماهنگ

است. این مؤلفان دریافتند که روش Tobit از دقت و ثبات بیشتری برخوردار است و واریانس بیشتری را تبیین می‌کند. بنابراین، استفاده از این مدل برای توزیعهای خاص بیش از مدل OLS کارآیی و دقت دارد و واریانس بیشتری از y را تبیین می‌کند. گروهی از صاحب‌نظران مانند جانستون و دناردو (۱۹۹۷) نیز به این نتیجه رسیده‌اند که در برخی از متغیرها از جمله غیبت کارکنان ـ که متغیر پیش‌بینی شونده منفی نمی‌شوند یعنی ($prob(y^*>0)$ است ـ می‌توان با استفاده از OLS به برآورد باثباتی دست یافت به شرط آنکه نسبت $\frac{n_1}{N}$ (یعنی نسبت مشاهده‌های سانسور شده به تعداد کل مشاهده‌ها) در β حاصل از مدل ضرب شود و از این طریق اریب از میان برود. اما آنان تأکید می‌کنند که ثبات این برآورد برای مواردی که در آن y^* یا x به طور پیوسته از توزیع نرمال برخوردار نیستند تضمین نمی‌شود. از آنجایی که غیبت کارکنان نیز دارای توزیع نرمال پیوسته نیست بنابراین استفاده از تصحیح نیز نمی‌تواند مفید باشد در نتیجه استفاده از مدل Tobit در مقایسه با مدل OLS در این مواقع ارجحیت دارد. بر این اساس، می‌توان پیشنهاد کرد که برای پیش‌بینی طول مدت غیبت کارکنان از سایر مدلها مانند رگرسیون OLS با داده‌های تبدیل شده، رگرسیون پواسون و رگرسیون لوجستیک استفاده و به مقایسه نتایج آنها پرداخته شود.

منابع

احمدیان، ع. (۱۳۷۶). *بررسی و ارزیابی عوامل مؤثر بر رضایت شغلی کارکنان اداره امور اقتصادی و دارایی شهرستان کاشان.* پایان‌نامه کارشناسی ارشد، دانشکده روان‌شناسی، دانشگاه آزاد اسلامی واحد خوراسگان اصفهان.

سرمد، ز.، بازرگان، ع. و حجازی، ا. (۱۳۷۶). *روشهای تحقیق در علوم رفتاری.* تهران :انتشارات آگاه.

فراهانی، ح. ا. (۱۳۸۲). *معضل هم‌خطی چندگانه در پژوهشهای رفتاری و راه‌حل آن.* هفته پژوهش، آذرماه ۱۳۸۲، دانشگاه اصفهان.

فراهانی، ح. ا. و عریضی، ح. ر. (۱۳۸۴). *روشهای پیشرفته پژوهش در علوم انسانی.* اصفهان: انتشارات جهاد دانشگاهی دانشگاه اصفهان.

کرلینگر، ان. و پدهازر، ا. (۱۳۸۴). رگرسیون چند متغیری در پژوهش رفتاری، ترجمه ح. سرایی. تهران : انتشارات سمت (تاریخ انتشار اثراصلی، ۲۰۰۱).

موسوی، م. (۱۳۸۴). *رابطه ناخشنودی شغلی با رفتارهای ضد تولید در سازمانهای تولیدی خدماتی شهر اصفهان.* پایان‌نامه کارشناسی ارشد، دانشکده روان‌شناسی، دانشگاه آزاد اسلامی واحد خوراسگان اصفهان.

وی‌ان. آ. و اسکاستر، ا. (۱۳۸۱). *تحلیل رگرسیون در علوم اجتماعی*، مترجم ح. نیرومند. مشهد: دانشگاه فردوسی مشهد (تاریخ انتشار اثراصلی، ۱۹۹۸).

هومن، ح. ع. (۱۳۸۰). *تحلیل داده‌های چندمتغیری در پژوهشهای رفتاری.* تهران : انتشارات پارسا.

Anostasopaylos, P. A., Troko, A. P., & Mamering, F. L. (2007). Tobit analysis of vehicle accident rates on interstate high ways. *Accident Analysis and Prevention* (in press).

Arbous, A. G., & Sichel, H. S. (1954). New techniques for the analysis of absenteeism data. *Biometrica, 41*, 77-90.

Baba, V. V. (1990). Methodological issues in modeling absence. *Journal of Applied Psychology, 75*, 428-432.

Geeg, C. W. (1983). Psychology of employee, lateness, absence and turnover : A methodological critique and empirical study. *Journal of Applied Psychology, 63*, 88-151.

Gibson, R. O. (1996). Toward a conceptualization of absence behavior. *Administrative Science Quarterly, 11*, 107-133.

Greene, W. H. (1993). *Econometric Analysis*. New York: Macmillan Publishing.

Hammer, T. H., & Landao, J. (1981). Methodological issues in the use of absence data. *Journal of Applied Psychology, 60*, 574-581.

Johnston, J., & Dinardo, J. (1997). Econometric methods (fourth ed.). USA: McGraw Hill.

Kuhn, T. S. (1970). *The structure of scientific revo-*

Madala, G. S. (1984). *Limited–dependent and qualitative variables in econometrics*. England : Cambridge University Press.

Price, J. L., & Mueller, C. W. (1986). *Absenteeism and turn over of hospital employees*. Greenwich, CT: JAI press.

Thoits, P. A. (1981). Undesirable life events and psycho-physiological distress. *American Sociological Review, 46*, 97-109.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica, 26*, 27-36.

Wang, L. (2007). A simple nonparametric test for diagnosing non linearity in Tobit median regression model. *Statistics & Probability, 77*, 1034-1042.

lutions. (second ed.) Chicago, IL: university of Chicago Press.

Landao, F. J., Vase, J. J., & Smith, F. D. (1984). *Methodological problems and strategies in predicting absence*. San Francisco, CA: Jossey–Bass.

Leigh, J. P. (1985). The effect of unemployment and the business cycle on absenteeism. *Journal of Economics and Business, 37* , 159-170.

Leigh, J. P. (2004). Hypertension , gender, job hazards and absenteeism in a 1973 national sample of N. S. Workers. *Journal of Clinical Epidemiology*, 293-314.

Lodahl, T. M., & Kejner, M. (1965). The definition and measurement of job involvement. *Journal of Applied Psychology, 49*, 24 -33.