

به کارگیری رگرسیون پواسنی و رگرسیون دوجمله‌ای منفی با انباشتگی صفر برای مدل سازی داده‌های آموزشی

***دکتر مسعود رودباری:** دانشیار، گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی ایران و مرکز تحقیقات مقاومت های میکروبی، دانشگاه علوم پزشکی ایران، تهران، ایران (نویسنده مسئول). m-roudbari@yahoo.co.uk

دکتر مسعود صالحی: استادیار، مرکز تحقیقات علوم مدیریت و اقتصاد سلامت، گروه آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی ایران، تهران، ایران. salehi74@yahoo.com

تاریخ پذیرش: ۹۲/۷/۲۸

تاریخ دریافت: ۹۲/۶/۲۵

چکیده

زمینه و هدف: بررسی تعداد دروس مردودی و ترم‌های مشروطی دانشجویان از جمله روش‌هایی است که می‌تواند موفقیت تحصیلی آنان را ارزیابی کند. از آن جا که این عوامل در بسیاری از دانشجویان مقدار صفر دارند، استفاده از مدل‌های رگرسیون پواسنی و دوجمله‌ای منفی با انباشتگی در صفر که برای مدل سازی داده‌های شمارشی و برآورد پارامترهای مدل بکار می‌رود، توصیه می‌گردد.

روش کار: در این مطالعه مقطعی که در سال تحصیلی ۸۸-۸۹ در دانشگاه علوم پزشکی ایران با جامعه آماری بیش از ۶۰۰۰ دانشجو و با نمونه‌ای ۶۷۰ نفری از آنان انجام شد، اطلاعات آموزشی، دانشجویی و جمعیت شناختی دانشجویان از طریق سایت سما و حوزه‌های مرتبط دانشگاه، اخذ گردید. نمونه گیری به صورت طبقه‌ای انجام پذیرفت. این مطالعه در شورای پژوهشی دانشگاه به تصویب رسید و کلیه اطلاعات دانشجویان به صورت محرمانه بررسی شد. برای تجزیه و تحلیل از نرم افزارهای SPSS و Stata استفاده شد.

یافته‌ها: در رگرسیون‌های پواسنی و دوجمله‌ای منفی با انباشتگی در صفر و برای متغیر وابسته تعداد ترم‌های مشروطی، معدل دانشجویان در دانشگاه و سهمیه بیش ترین نقش را در مدل داشته‌اند. در مدل های فوق و برای متغیر وابسته تعداد دروس مردودی، معدل دانشگاه، مقاطع کارشناسی و کارشناسی ارشد در هر دو مدل حضور داشته و مدل‌ها شامل برخی دیگر از عوامل موثر در تعداد دروس مردودی نیز بوده‌اند.

نتیجه گیری: در رگرسیون‌های تعداد ترم‌های مشروطی، معدل و سهمیه نقش مهمی دارند به طوری که افزایش معدل باعث کاهش تعداد ترم‌های مشروطی می‌شود. همچنین تغییر از سایر سهمیه‌ها به سهمیه مناطق (آزاد) باعث کاهش تعداد ترم‌های مشروطی می‌گردد. در رگرسیون‌های تعداد دروس مردودی، معدل در هر دو مدل نقش مهمی دارد و افزایش آن باعث کاهش تعداد دروس مردودی می‌شود. همچنین مقاطع کارشناسی و و کارشناسی ارشد نیز باعث کاهش تعداد دروس مردودی می‌گردند.

کلیدواژه‌ها: پواسن، دو جمله ای منفی، انباشتگی در صفر، مردودی، مشروطی، دانشجو، ایران

مقدمه

در آمار زیستی برای بررسی متغیرهای گسسته (شمارشی) معمولاً از توزیع پواسن (اگر میانگین و واریانس توزیع مساوی باشند) و از توزیع دوجمله‌ای منفی (اگر واریانس توزیع از میانگین آن بیشتر باشد به عبارت دیگر بیش پراکنش) استفاده می‌شود. در شرایطی که هدف بررسی، توصیف و تبیین یک متغیر شمارشی با استفاده از چند متغیر باشد، معمولاً از مدل سازی آماری استفاده می‌شود.

مدل سازی آماری از جمله روش‌های مناسب برای بررسی روابط بین پدیده‌های گوناگون به خصوص پدیده‌های پزشکی و بهداشتی می‌باشند. یکی از مدل‌هایی که برای مدل سازی پدیده‌های

شمارشی نظیر تعداد دفعات اهدای خون، تعداد دفعات ترک اعتیاد، تعداد دروس مردودی و ... به کار می‌رود و در بسیاری از مطالعات پزشکی کاربرد دارد، رگرسیون‌های پواسنی است. در این روش به مدل سازی میانگین این متغیرها با استفاده از روابط آماری پرداخته می‌شود. رگرسیون‌های پواسنی زیرمجموعه‌ای از مجموعه‌ای بزرگ از مدل‌های آماری هستند که به مدل‌های تعمیم یافته خطی (Generalized Linear Model = GLM) مشهور هستند.

گاهی اوقات متغیرهای شمارشی که از آن‌ها برای مدل سازی آماری استفاده می‌شود، دارای انباشتگی بیش از حد در صفر هستند (فراوانی بسیار بزرگ در صفر) که با توجه به توزیع متغیر

دارای انباشتگی زیاد در نقطه صفر و تعداد ترم‌های مشروطی نیز دارای انباشتگی بسیار زیاد در نقطه صفر می باشد، استفاده از رگرسیون پواسنی با انباشتگی در صفر و یا رگرسیون دوجمله‌ای منفی با انباشتگی در صفر می باشد، کارآیی زیادی دارند. هدف این مطالعه مدل سازی تعداد دروس مرودی و ترم‌های مشروطی دانشجویان دانشگاه علوم پزشکی ایران با استفاده از توزیع پواسن با انباشتگی صفر و یا توزیع دوجمله‌ای منفی با انباشتگی صفر و تعیین متغیرهای تاثیر گذار بر تعداد دروس مرودی و ترم‌های مشروطی آنان می باشد.

روش کار

این مطالعه از نوع مقطعی بوده که در سال‌های ۱۳۸۸ الی ۱۳۸۹ در دانشگاه علوم پزشکی ایران انجام شد. جامعه آماری مطالعه کلیه دانشجویان شاغل به تحصیل دانشگاه علوم پزشکی ایران در سال تحصیلی ۸۹-۸۸ بود که بیش از ۶۰۰۰ نفر در مقاطع مختلف تحصیلی بوده‌اند. با استفاده از نمونه گیری طبقه‌ای که در آن طبقه‌ها دانشکده‌های دانشگاه هستند، نمونه گیری صورت گرفته که حجم نمونه ۶۷۰ نفر محاسبه شد.

داده‌های دموگرافیک و آموزشی از طریق سایت سما و اطلاعات مربوط به حضور دانشجویان در خوابگاه دانشگاه از طریق حوزه دانشجویی دانشگاه جمع‌آوری شد. این اطلاعات شامل اسامی، شماره دانشجویی دانشجویان، رشته، دانشکده (یک برای هر دانشکده و صفر برای سایر دانشکده‌ها)، مقطع تحصیلی (یک برای هر مقطع و صفر برای سایر مقاطع)، جنسیت (مرد=۱ و زن=صفر)، سال ورود، وضعیت تاهل (بر اساس اطلاعات زمان ورود، مجرد=۱ و متاهل=صفر)، استفاده از خوابگاه، بومی بودن، سهمیه (سهمیه آزاد=۱ و سایر سهمیه‌ها=صفر)، معدل دانشگاه، تعداد دروس مرودی و تعداد ترم‌های مشروطی بود.

برای انتخاب نمونه‌ها از فایل موجود، به هر دانشجو یک کد داده شده و با روش تصادفی ساده در نرم افزار Minitab نمونه مورد نظر انتخاب شده و فایل موجود از فایل اصلی داده‌ها جدا گردید.

مورد مطالعه، به توزیع پواسنی با انباشتگی صفر و یا توزیع دوجمله‌ای منفی با انباشتگی صفر تقسیم می‌شوند.

رفیعی و همکاران در مطالعه‌ای به منظور تحلیل مدت زمان اقامت مادران پس از زایمان از توزیع‌های مختلف گاوسین، گاوسین معکوس، پواسن، دوجمله‌ای منفی، گاما، دوجمله‌ای منفی با تابع واریانس دوجمله‌ای منفی، مدل‌های بریده شده پواسن و دوجمله‌ای منفی و با تابع پیوند همانی و لگاریتمی استفاده کردند که مدل دوجمله‌ای منفی و دو جمله‌ای منفی بریده شده بهترین مدل برای مدل سازی مدت زمان اقامت مادران در بیمارستان می باشد (۱).

وونک و همکاران در مطالعه‌ای در سال ۲۰۱۳ بر روی DMF دانش آموزان هنک کنگی که از رگرسیون پواسنی با انباشتگی در صفر استفاده شد، به بررسی وضعیت سلامتی این دانش آموزان پرداختند (۲).

براندس و همکاران در مطالعه‌ای در آمریکا در سال ۲۰۱۰ برای مدل سازی افرادی که اخیراً به سیگار کشیدن روی آورده‌اند از رگرسیون پواسنی با انباشتگی در صفر استفاده کرده و تعداد دفعات سیگار کشیدن آن‌ها را با توجه به متغیرهای گوناگون از جمله نژاد افراد سیگاری، بررسی کردند (۳).

محمدفام و همکاران در مطالعه سال ۲۰۰۹ در همدان از رگرسیون پواسنی با انباشتگی در صفر برای مدل سازی حوادث حین کار استفاده کردند. آنها با استفاده از این مدل نشان دادند که وضعیت تاهل و مقیاس صدمات در تعداد حوادث حین کار نقش دارد. همچنین تکنسین‌ها و سوپروایزرها بیشتر از کارگراها در معرض حوادث ضمن کار بوده‌اند (۴).

از جمله مواردی که این روش‌ها کارآیی دارد محاسبه تعداد دروس مرودی و ترم‌های مشروطی دانشجویان با توجه به خصوصیات دانشجویان و برآورد تعداد دروس مرودی یا ترم‌های مشروطی با توجه به متغیرهای دموگرافیک و آموزشی آنان می باشد. در این قبیل بررسی‌ها با توجه به این که تعداد دروس مرودی

یک یا چند درس مرودی داشتند. تعداد دانشجویان با یک، دو، سه، چهار یا بیشتر از ۴ درس مرودی به ترتیب ۹۲ نفر (۱۳/۷٪)، ۳۷ نفر (۵/۵٪)، ۲۲ نفر (۳/۳٪) و ۴۱ نفر (۶/۲٪) بودند. میانگین تعداد دروس مرودی ۰/۸۳ (با انحراف معیار ۲/۲) بوده و حداکثر دروس مرودی ۱۹ بود. تعداد دانشجویان بدون ترم مشروطی ۶۱۵ نفر (۹۱/۸٪) و ۵۵ نفر (۸/۲٪) نیز ترم مشروطی داشتند. همچنین تعداد ۲۴ نفر (۳/۶٪) تنها یک ترم مشروطی داشتند و تعداد افراد با دو یا بیش از دو ترم مشروطی ۳۱ نفر (۴/۶٪) بودند. میانگین تعداد ترم های مشروطی ۰/۱۶ (با انحراف معیار ۰/۶۳۵۲) و حداکثر تعداد ترم های مشروطی ۵ بود.

نتایج حاصل از برازش رگرسیون پواسنی با انباشتگی در صفر در جدول ۱ ارائه شده است. همان طور که در جدول ۱ می توان دید، در رگرسیون تعداد دروس مرودی، افزایش یک واحدی معدل دانشگاه باعث کاهش ۰/۶۴ لگاریتم تعداد دروس مرودی می شود. با افزایش یک واحدی تعداد ترم های مشروطی، لگاریتم بخت صفر انباشته به اندازه ۲/۷۳ کاهش می یابد. تغییر مقطع تحصیلی از سایر مقاطع به کارشناسی یا کارشناسی ارشد باعث کاهش لگاریتم مورد انتظار تعداد دروس مرودی به ترتیب به اندازه ۰/۲۷ و ۰/۹۶ می شود. همچنین تغییر جنسیت از

فایل جدا شده از اطلاعات کلیه دانشجویان دانشگاه به صورت فایل Stata ذخیره شده و در تجزیه و تحلیل ها بکار گرفته شد. افراد انتخاب شده در دانشکده ها براساس تعداد دانشجویان و مقطع تقسیم بندی شده و از هر مقطع به نسبت آن دانشجو انتخاب شده است. برای الحاق فایل اطلاعات دانشجویان با فایل تعداد دروس مرودی و ترم های مشروطی که فایل دوم از نظر تعداد رکورد بسیار کوچک تر از فایل اول بوده است با جستجوی شماره دانشجویان در فایل مرودین و مشروطین و اضافه کردن آن به فایل اطلاعات اصلی، فایل اطلاعات تکمیل گردید.

برای تجزیه و تحلیل داده ها از نرم افزار STATA نسخه ۹/۱ و روش های آمار توصیفی و رگرسیون پواسنی با انباشتگی در صفر (Zero Inflated Poisson Regression: ZIP) و رگرسیون دو جمله ای منفی با انباشتگی در صفر (Zero Inflated Negative Regression: ZINB) استفاده شد. کلیه اطلاعات دانشجویان محرمانه بوده و به صورت کدبندی شده مورد تجزیه و تحلیل قرار گرفت. همچنین طرح در شورای پژوهشی دانشگاه مورد تصویب قرار گرفت.

یافته ها

تعداد دانشجویان که درس مرودی نداشتند ۴۷۸ نفر (۷۱/۳٪) بوده و ۱۹۲ نفر (۲۸/۷٪) نیز در

جدول ۱- نتایج حاصل از برازش مدل های رگرسیون پواسنی با انباشتگی در صفر

رگرسیون با متغیر پاسخ تعداد ترم های مشروطی			متغیر	رگرسیون با متغیر پاسخ تعداد دروس مرودی			متغیر
برآورد	خطای معیار	مقدار احتمال		برآورد	خطای معیار	مقدار احتمال	
-۰/۴۲	۰/۰۸۸۴	<۰/۰۰۰۱	معدل دانشگاه	-۰/۶۴	۰/۰۳۳۹	<۰/۰۰۰۱	معدل دانشگاه
-۰/۴۷	۰/۲۰۹۲	۰/۰۲۳	سهامیه	-----	-----	-----	سهامیه
-----	-----	-----	مقطع کارشناسی ارشد	-۰/۹۶	۰/۳۷۵۳	۰/۰۱۱	مقطع کارشناسی ارشد
-----	-----	-----	مقطع کارشناسی	-۰/۲۷	۰/۰۹۸۳	۰/۰۰۵	مقطع کارشناسی
-----	-----	-----	جنسیت	-۰/۲۴	۰/۰۹۵۹	۰/۰۱۳	جنسیت
-----	-----	-----	تاهل	۰/۰۵	۰/۰۱۹۷	۰/۰۱۶	تاهل
انباشتگی در صفر (تابع ربط لوجیت)							
-۱/۵۶	۰/۲۸۲۷	<۰/۰۰۰۱	تعداد دروس مرودی (انباشتگی در صفر)	-۲/۷۳	۱/۱۸۰۵	۰/۰۲۱	تعداد ترم های مشروطی (انباشتگی در صفر)

جدول ۲- نتایج حاصل از برازش مدل‌های رگرسیون دو جمله‌ای منفی با انباشتگی در صفر

رگرسیون با متغیر پاسخ تعداد ترم‌های مشروطی				رگرسیون با متغیر پاسخ تعداد دروس مردودی*		
مقدار احتمال	خطای معیار	برآورد	متغیر	مقدار احتمال	خطای معیار	برآورد
<0/0001	0/0884	-0/42	معدل دانشگاه	<0/0001	0/0515	-0/8350
0/023	0/2092	-0/47	سه‌میه	-----	-----	-----
-----	-----	-----	دانشکده مدیریت	<0/0001	0/3273	1/24
-----	-----	-----	دانشکده توانبخشی	0/001	0/3239	1/07
-----	-----	-----	مقطع کارشناسی	0/002	0/2636	-0/8098
-----	-----	-----	مقطع کاردانی	0/002	0/2854	-0/8847
-----	-----	-----	مقطع کارشناسی ارشد	0/004	0/5025	-1/4545
<0/0001	0/2827	-1/56	تعداد دروس مردودی	0/021	1/1805	-35/88

* - با استفاده از روش Robust

کارشناسی، کاردانی و کارشناسی ارشد، لگاریتم مورد انتظار تعداد دروس مردودی به ترتیب به اندازه 0/8098، 0/8847 و 1/4545 کاهش می‌یابد.

در رگرسیون تعداد ترم‌های مشروطی، متغیر معدل با تعداد ترم‌های مشروطی رابطه معکوس دارد به طوری که با افزایش یک واحدی معدل، لگاریتم مورد انتظار تعداد ترم‌های مشروطی 0/42 کاهش می‌یابد. همچنین با تغییر سه‌میه از سایر سه‌میه‌ها به سه‌میه مناطق، لگاریتم مورد انتظار تعداد ترم‌های مشروطی به اندازه 0/47 کاهش می‌یابد. علاوه بر این با افزایش یک واحدی تعداد دروس مردودی، لگاریتم بخت صفر انباشته به اندازه 1/56 کاهش می‌یابد که کلیه این نتایج کاملاً با نتایج جدول ۱ در مورد رگرسیون با متغیر پاسخ تعداد ترم‌های مشروطی یکسان می‌باشد.

بحث و نتیجه‌گیری

یافته‌های تحقیق نشان می‌دهند که با استفاده از رگرسیون پواسنی و دو جمله‌ای منفی با انباشتگی درصفر و با متغیر پاسخ تعداد ترم‌های مشروطی، متغیرهای معدل دانشگاه و سه‌میه ارتباط معکوسی با متغیر پاسخ دارند به طوری که افزایش معدل و تغییر از سایر سه‌میه‌ها به سه‌میه مناطق باعث کاهش ترم‌های مشروطی می‌شوند. در تحقیقی نشان داده شده است که دانشجویان

مرد به زن باعث کاهش لگاریتم مورد انتظار تعداد دروس مردودی به اندازه 0/24 و تغییر از وضعیت متاهل به مجرد باعث افزایش لگاریتم مورد انتظار تعداد دروس مردودی به اندازه 0/05 می‌شود.

در رگرسیون تعداد ترم‌های مشروطی، متغیر معدل با تعداد ترم‌های مشروطی رابطه معکوس دارد به طوری که با افزایش یک واحدی معدل، لگاریتم مورد انتظار تعداد ترم‌های مشروطی 0/42 کاهش می‌یابد. همچنین با تغییر سه‌میه از سایر سه‌میه‌ها به سه‌میه مناطق، لگاریتم مورد انتظار تعداد ترم‌های مشروطی به اندازه 0/47 کاهش می‌یابد. علاوه بر این با افزایش یک واحدی تعداد دروس مردودی، لگاریتم بخت صفر انباشته به اندازه 1/56 کاهش می‌یابد. نتایج برازش رگرسیون دو جمله‌ای منفی در جدول ۲ ارائه شده است.

همان‌طور که مشاهده می‌شود در رگرسیون تعداد دروس مردودی، افزایش یک واحدی معدل دانشگاه باعث کاهش 0/83 لگاریتم تعداد دروس مردودی می‌شود. با افزایش یک واحدی تعداد ترم‌های مشروطی، لگاریتم بخت صفر انباشته به اندازه 35/88 کاهش می‌یابد. تغییر سایر دانشکده‌ها به دانشکده مدیریت و توانبخشی، باعث افزایش لگاریتم مورد انتظار تعداد دروس مردودی به ترتیب به اندازه 1/24 و 1/07 می‌شود. با تغییر مقطع تحصیلی از سایر مقاطع به مقاطع

دروس مردودی، متغیرهای جنسیت و تاهل به ترتیب تاثیر معکوس و مستقیم بر تعداد دروس مردودی دارند به طوری که تعداد دروس مردود در مردان بیشتر از زنان و در مجردین بیشتر از متاهلین است. در مطالعات مختلفی (۵ و ۶). نشان داده شده است که جنس و تاهل در تعداد دروس مردودی یا ترم‌های مشروطی دانشجویان تاثیر دارد که این نتیجه با نتیجه حاصل از این تحقیق همسو است. در تحقیق یوسفی مشعوف و همکاران نشان داده شده است که مجردین بیشتر از متاهلین در تحصیل موفق هستند و تعداد دروس مردودی کمتری دارند که این نتیجه با نتایج این تحقیق مغایرت دارد (۷).

در مدل رگرسیون دو جمله ای منفی، با متغیر پاسخ تعداد دروس مردودی، متغیرهای مقطع کارشناسی و دانشکده‌های مدیریت و توانبخشی باعث افزایش تعداد دروس مردودی می‌شوند. با توجه به اینکه تعداد دروس مردودی در مقطع کارشناسی بیشتر از سایر مقاطع می‌باشد، نتیجه بدست آمده مورد انتظار بوده است. همچنین دانشکده توانبخشی بیش‌ترین میزان مردودی را در دانشگاه داشته است که هم راستا با نتایج مدل بوده است. به نظر می‌رسد ورود دانشکده مدیریت به مدل به عنوان دانشکده ای که مردودی زیادتری داشته است، به علت تعدیل اثرات متغیرهای مختلف در مدل می‌باشد.

از سوی دیگر افزایش یک واحدی تعداد ترم‌های مشروطی در هر دو مدل، لگاریتم بخت صفر انباشته را نیز کاهش می‌دهد که این کاهش در مدل رگرسیون دو جمله ای منفی بیش از ۱۳ برابر مدل رگرسیون پواسنی می‌باشد که این نتایج با وجود همبستگی خوب و مستقیم تعداد دروس مردودی و تعداد ترم‌های مشروطی مورد انتظار بوده است.

از نتایج این طور نتیجه گیری می‌شود که در رگرسیون‌های تعداد ترم‌های مشروطی، معدل و سهمیه نقش مهمی را ایفا می‌کنند به طوری که افزایش معدل باعث کاهش تعداد ترم‌های مشروطی می‌شود. همچنین تغییر از سایر سهمیه‌ها به سهمیه مناطق (آزاد) نیز باعث کاهش تعداد

سهمیه مناطق (آزاد) تعداد ترم‌های مشروطی کمتری نسبت به سایر افراد دارد که با نتیجه این تحقیق یکسان است (۵). این نتیجه با این واقعیت که پدیده مشروطی در دانشجویان ممتاز و با معدل‌های بالا بسیار نادر بوده و این دانشجویان کمتر در معرض مشروط شدن قرار می‌گیرند، همسان است. همچنین دانشجویان سهمیه مناطق نیز اکثراً از دانشجویان ممتاز بوده و پدیده مشروطی در بین این گروه بسیار نادر است، بنابراین سهمیه مناطق از عوامل کاهش تعداد ترم‌های مشروطی دانشجویان می‌باشد. از سوی دیگر افزایش یک واحدی تعداد دروس مردودی در هر دو مدل، لگاریتم بخت صفر انباشته را نیز کاهش می‌دهد. بنابراین افزایش تعداد دروس مردودی باعث افزایش تعداد ترم‌های مشروطی می‌شود. این مساله با توجه به همبستگی نسبتاً بالای (۰/۸۱) تعداد ترم‌های مشروطی و دروس مردودی، که با افزایش تعداد دروس مردودی شانس مشروط شدن دانشجویان نیز افزایش می‌یابد، همسو می‌باشد.

در هر دو مدل رگرسیونی با متغیر وابسته تعداد دروس مردودی، معدل دانشگاه و مقاطع کارشناسی و کارشناسی ارشد رابطه معکوس با متغیر پاسخ دارند، به طوری که در هر دو مدل افزایش معدل باعث کاهش تعداد دروس مردودی می‌شود که میزان این کاهش در رگرسیون دو جمله ای منفی بیشتر از رگرسیون پواسنی است. در تحقیقی نشان داده شده است که بین معدل دانشگاه و تعداد دروس مردودی رابطه معکوس موجود است که با نتیجه تحقیق جاری همخوانی دارد (۵). همچنین تغییر از سایر مقاطع تحصیلی به مقاطع کارشناسی و کارشناسی ارشد نیز باعث کاهش تعداد دروس مردودی در این مقاطع تحصیلی می‌شود که این کاهش‌ها در مدل رگرسیون دو جمله ای منفی بیشتر از مدل رگرسیون پواسنی می‌باشد. کاهش تعداد دروس مردودی در مقاطع کارشناسی و کارشناسی ارشد با توجه به کم بودن تعداد دروس مردودی در این دو مقطع، منطقی می‌باشد.

در مدل رگرسیونی پواسنی با متغیر پاسخ تعداد

41.

4. Mohammadfam I, Moghimbeigi A. Evaluation of injuries among a manufacturing industry staff in Iran. *J Res Health Sci*. 2009 Jun 28; 9(1):7-12.

5. Roudbari M, Ahmadi A, Ebadifard Azar F. Determination of the effective factors on the educational progress of the students of Tehran University of medical education in 88-89. *Teb va Tazkieh*. 1391; 19(3):37-48.

6. Shams B, Farshidfar M, and Hassanzadeh A. The comparison of demographic and personality characteristics in fail and succeed students of Isfahan University of medical sciences. *Research in Medical Science*, 1997; 4(2): 222-226.

7. Yousefi Mashoof, R. and M. Saeedi Jam, Study in quality of education status of medical students in basic sciences courses in Hamadan university of medical sciences 1989-94. *The scientific Journal of Hamadan University of Medical Sciences*, 2001; 7(4): 25-29.

ترم‌های مشروطی می‌گردد. علاوه بر این در این مدل‌ها افزایش تعداد دروس مرودی نیز باعث کاهش لگاریتم بخت صفر انباشته می‌گردد. بنابراین با افزایش معدل دانشجویان و انتخاب دانشجویان سهمیه آزاد می‌توان تعداد ترم‌های مشروطی را کاهش داد.

در رگرسیون‌های تعداد دروس مرودی، معدل در هر دو مدل نقش مهمی دارد و افزایش آن باعث کاهش تعداد دروس مرودی می‌شود. همچنین مقاطع کارشناسی و و کارشناسی ارشد نیز باعث کاهش تعداد دروس مرودی می‌گردند. سایر متغیرهای تاثیر گذار بر تعداد دروس مرودی در دو مدل متفاوت است. از سوی دیگر افزایش تعداد ترم‌های مشروطی، لگاریتم صفر انباشته را در هر دو مدل کاهش می‌دهد که میزان آن در دو مدل متفاوت است. بنابراین افزایش معدل دانشجویان و انتخاب بیشتر دانشجویان کارشناسی و کارشناسی ارشد، می‌تواند تعداد دروس مرودی را کاهش دهد.

تقدیر و تشکر

این مقاله حاصل طرح تحقیقاتی در دانشگاه علوم پزشکی ایران با شماره ۳۷۰ بوده است. نویسندگان بر خود لازم می‌دانند که از مرکز مطالعات و توسعه آموزش دانشگاه به خاطر حمایت‌های مالی و از مدیریت تحصیلات تکمیلی و مدیریت امور دانشجویی دانشگاه به خاطر تحویل اطلاعات آموزشی و دانشجویی دانشجویان دانشگاه تشکر و قدردانی نمایند.

منابع

1. Rafiee M, Ayatollahi M, Behboodjan J. Zero-inflated negative binomial modeling, efficiency for analysis of length of maternity hospitalization. *Yafteh*. 2005; 6 (4):47-58.

2. Wong KY, Lam KF. Modeling zero-inflated count data using a covariate-dependent random effect model. *Stat Med*. 2013 Apr 15;32(8):1283-93.

3. Barondess DA, Meyer EM, Boinapally PM, Fairman B, Anthony JC. Epidemiological evidence on count processes in the formation of tobacco dependence. *Nicotine Tob Res*. 2010 Jul;12(7):734-

The Poisson and negative binomial regression models with zero-inflated count: An application to educational data

Masoud Roudbari, Associate Professor, the Anti-Microbial Resistance Research Centre, Department of Biostatistics, Faculty of Public Health, Iran University of Medical Sciences, Tehran, Iran (*Corresponding author). mroudbari@yahoo.co.uk

Masoud Salehi, Assistant Professor of Biostatistics, Member of Health Management and Economics Research Center, Department of Biostatistics, Faculty of Public Health, Iran University of Medical Sciences, Tehran, Iran. Salehi74@yahoo.com

Abstract

Background: Investigation through the number of failed courses and semesters in medical students is a method for assessing students' performance. Since these amounts have a zero inflated (ZI) value in most students, use of Poisson and negative binomial distributions with ZI, in which can model count data to find the associated factors and estimate the parameters, is recommended.

Methods: This cross-sectional study was performed between 2008 and 2009 in Iran University of Medical Sciences (IUMS) on 670 sample students who were selected using stratified sampling. The educational and demographic data was collected from University site and related resources. The study was approved in the IUMS Research Council and Board of Ethics and accordingly the students' data was investigated confidentially. For data analysis SPSS and Stata softwares were used.

Results: Up to Poisson regression and negative binomial distributions with ZI for dependent variables of the number of failed semesters, students' total averages and the enrolment quota had the most roles in the model. Dependent variables of the number of failed course, total average and undergraduate and master levels had the most effect in both models.

Conclusion: In ZI regression model for the number of failed semester, the total average and enrolment quota type have important roles so that increasing the average yields the decreasing of the number of failed semester. Moreover, moving from the other quota types to open-for-all enrolment quota yields a decrease in the number of failed semesters. In regressions with the number of failed courses, the average has an important effect to both models and its increase yields a decrease of the number of failed courses. Also, the undergraduate and master levels cause a decrease in the number of failed courses.

Keywords: Poisson, Negative binomial, Zero inflated, Failing, Iran