



کنگره علوم و مهندسی آب و فاضلاب ایران

دانشگاه تهران، تهران

۲۶ و ۲۷ بهمن ماه ۱۳۹۵

10520-NWWCE

پیش‌بینی نرخ آسیب‌پذیری لوله‌های شبکه آب شرب شهری با بکارگیری و توسعه مدل‌های آماری

سونیا قاسم نژاد^۱، همایون مطیعی^۲، سعید موسوی ندوشنی^۳

۱- دانش‌آموخته کارشناسی ارشد، دانشکده آب و محیط‌زیست، دانشگاه شهید بهشتی

۲ و ۳- عضو هیئت علمی، دانشکده آب و محیط‌زیست، دانشگاه شهید بهشتی

H_motiei@sbu.ac.ir

خلاصه

شکست لوله‌ها در شبکه‌های توزیع آب شهری، باعث نشت جریان از شبکه شده و نه تنها باعث هدر رفت مقادیر قابل توجهی از آب تصفیه‌شده می‌گردد، سبب اتلاف سرمایه‌های مادی و خدمات انسانی پشتیبان نیز می‌شود. مهم‌ترین متغیرهای شکست لوله‌ها عبارتند از جنس، سن، طول، قطر و در این مقاله، از چهار روش آماری، جهت یافتن معادلاتی که بتوانند به تخمین احتمال شکست لوله‌ها در شبکه آب شرب شهری و تعیین پارامترهایی که بیشترین تأثیر را در شکست دارند، استفاده شده است. با تحلیل داده‌های موجود و به کمک روش‌های رگرسیونی، رابطه‌ی میان پارامترهای تأثیرگذار و نرخ وقوع حوادث به وقوع پیوسته، به دست آمده است. این چهار مدل آماری عبارتند از: مدل رگرسیون خطی، مدل رگرسیون نمایی، مدل رگرسیون پواسون و مدل رگرسیون لژیستیک. به منظور ارزیابی روش‌های ارائه شده از داده‌های جمع‌آوری شده حوادث لوله‌ها در شبکه‌ی توزیع آب ناحیه‌ی ۱ از منطقه یک آب و فاضلاب شهر تهران، استفاده گردید و نتایج نشان دادند که از میان مدل‌های آماری بررسی شده، مدل رگرسیون لژیستیک قادر است وقایع آینده را با احتمال بهتری پیش‌بینی کند.

کلمات کلیدی: شبکه‌های آب شهری، مدل‌های آماری، رگرسیون، شکست لوله، حوادث و اتفاقات.

۱. مقدمه

خرابی و شکست لوله‌ها در سیستم توزیع آب، چالش‌های زیادی را برای شرکت‌های تأمین‌کننده‌ی آب در سراسر جهان ایجاد می‌کند. پوسیدگی لوله‌ها منجر به شکست لوله‌ها و نشت در آن‌ها می‌گردد که این باعث کاهش ظرفیت حمل آب در لوله‌ها و افزایش هزینه‌های تعمیرات می‌گردد. شکست لوله‌ها همچنین منجر به کاهش ظرفیت آتش‌نشانی و آلودگی آب سیستم می‌گردد. تولید آب سالم و بهداشتی در شهرهای مختلف کشور و به‌خصوص تهران با زحمت فراوان و طی مراحل مختلف و صرف هزینه‌های هنگفت صورت می‌پذیرد، از استحصال آب از منابع و راه‌اندازی خطوط انتقال و ایستگاه‌های پمپاژ تا تصفیه‌ی آن و توزیع در شبکه و رساندن به دست مشترکین و اتلاف آب معادل است با افزایش در هزینه‌های تولید، انتقال، تصفیه، توزیع و هزینه‌های بهره‌برداری.

با توجه به نقش حیاتی آب در زندگی بشر، کمبود منابع آب قابل شرب و هزینه‌های گزاف تعمیر و بازسازی شبکه، یکی از فعالیت‌ها و تصمیم‌های مدیریتی، کاهش تعداد حوادث و اتفاقات سامانه‌های آب‌رسانی شهری است که این امر منجر به کاهش تلفات آب و همچنین کاهش هزینه‌های تعمیر و نگهداری و بازسازی می‌شود. به‌منظور دستیابی به اهداف مذکور لازم است بررسی کاملی از حوادث و اتفاقات شبکه، عوامل مؤثر در ایجاد حادثه و تأثیر مشخصات لوله‌ها در تعداد حوادث و اتفاقات شبکه، عوامل مؤثر در ایجاد حادثه و در سامانه‌های توزیع با دقت خوبی پیش‌بینی شود و جهت رفع آن اقدام لازم به عمل آید. در تأکید اهمیت موضوع می‌توان اعلام نمود که در سال ۱۳۷۷، حدود یک میلیون حادثه در سامانه‌های توزیع آب کشور رخ داده که بیش از ۲۰ درصد از کل درآمدهای شرکت آب و فاضلاب کشور را برای تعمیر، بازسازی و اصلاح به خود اختصاص داده است که حدود ۳۰ درصد این حوادث روی لوله‌های سامانه توزیع بوده است [1]. از طرفی دیگر، بررسی‌ها نشان می‌دهند که میزان هزینه‌های نگهداری

و تعمیرات سنتی در بخش آب، از ۹۴ میلیارد ریال در سال ۱۳۷۸ به ۲۳۹ میلیارد ریال در سال ۱۳۸۱، افزایش یافته است. [2] تاکنون روش های مختلفی برای مدل کردن و آنالیز شکست لوله ها، قابلیت اعتماد و عمر باقیمانده ی آنها ایجاد شده است:

(۱) مدل سازی فیزیکی

(۲) مدل سازی توصیفی

(۳) مدل سازی آماری

در تحلیل فیزیکی نتایجی قوی حاصل می گردد ولی به دلیل مدفون بودن لوله ها، جمع آوری اطلاعات مربوطه برای این مدل ها در سیستم توزیع آب کاری دشوار و بسیار هزینه و زمان بر است و از این جهت فقط برای لوله های انتقال اصلی مورد استفاده قرار می گیرد. آنالیز توصیفی شامل محاسبه ی آمار توصیفی به منظور ایجاد یک دید کلی راجع به الگوی شکست در لوله ها است. این روش در لوله هایی قابل استفاده است که داده های جامعی درباره ی ویژگی های لوله ها و شکست در آنها وجود دارد و از آنجایی که غالباً داده های در دسترس سیستم توزیع آب در این سطح نمی باشند، لذا از روش آماری برای پیش بینی رفتار لوله های خطوط توزیع در شکست استفاده می گردد زیرا استراتژی مدل کردن برای به حداقل رسانیدن داده های مورد نیاز و در نظر گرفتن این واقعیت است که تعداد شکست های قبلی یک عامل کلیدی در مدل سازی است. مدل های آماری دید کلی مناسبی درباره ی الگوی شکست لوله ها ایجاد کرده و به کمی سازی روند وقوع آنها با دقت بیشتر کمک می کند. اکثر قواعد موجود جهت ارزیابی عملکرد آینده ی لوله های در حال فرسودگی و اخذ تصمیماتی درباره ی تعویض یا تعمیرات آنها فقط بر اساس سن لوله و تعداد شکست های قبلی اند. مثال هایی از آنها شامل مدل خطی زمانی^۱ [3] مدل نمایی زمانی^۲ [4] می باشند. از این جهت مطالعاتی به منظور تکمیل و تأیید این روش ها مورد نیاز است. طبیعت پیچیده ی مکانیسم شکست و متغیر بودن نرخ شکست در لوله های هر شبکه و سیستم توزیع آب، باعث تلاش بیشتر برای دستیابی به مدل های پیش بینی کننده ی شکست های آینده است [5]. تحلیل های آماری شکست به منظور انجام بررسی و اخذ تصمیماتی راجع به لوله ها بسیار به صرفه است. هدف از این تحقیق، یافتن معادلاتی است که از مدل های آماری بدست آمده و بتوان از آنها برای شناخت الگوی شکست لوله ها استفاده کرد. مدل های به کار رفته در این تحقیق عبارت اند از مدل خطی، مدل نمایی، مدل خطی تعمیم یافته پواسون^۳ و مدل خطی تعمیم یافته لجستیک^۴. و متغیرهای به کار رفته عبارت اند از قطر لوله، طول لوله، جنس لوله، سن لوله و فشار داخلی لوله.

۲. مواد و روش ها

روش رگرسیون از جمله روش های آماری برای بررسی و مدل سازی ارتباط بین متغیرها است. رگرسیون تقریباً در هر زمینه ای از جمله مهندسی، فیزیک و غیره برای برآورد و پیش بینی مورد استفاده است. تحلیل رگرسیونی، پرکاربردترین روش در بین تکنیک های آماری است و آنچه در رگرسیون حائز اهمیت می باشد، این است که معادله ای که بیانگر رابطه ی بین متغیرهاست بدست آمده و از روی معادله ی به دست آمده، پیش بینی یا برآوردهای لازم را انجام گیرد. در این تحقیق از چهار مدل رگرسیونی به شرح زیر استفاده شده است:

۲-۱- مدل رگرسیون خطی

در این مدل فرض بر آن است که متغیر Y تابعی از متغیر توصیفی X_i است.

$$Y = \beta_0 + \sum_{i=1}^n \beta_i X_i + \varepsilon \quad (1)$$

در این معادله β_0 و β_i مقادیر ثابت اند (پارامترهای رگرسیون) که برآورد می شوند. ε مقدار خطا با فرض این که خطاها با میانگین صفر و واریانس نامعلوم، توزیع نرمال دارند و مستقل اند [6]. رابطه ی خطی بین تعداد شکست های یک بخش از لوله و سن لوله، اولین بار توسط Shamir and Howard [7] نظیر معادله ی (۲) مطرح گردید.

¹ Time linear model

² Time exponential model

³ Poisson generalized linear model

⁴ Logistic generalized linear model

$$N = k_0 \times A \quad (2)$$

در این معادله N تعداد شکست‌ها در سال در هر قطعه از لوله، k_0 پارامتر رگرسیون و A سن لوله در اولین شکست می‌باشد. این نتایج حاصل از بررسی داده‌های شکست در طی ۱۰ سال در شهر Winnipeg کانادا می‌باشند. آن‌ها همچنین دریافتند که رابطه‌ای خطی منفی بین قطر لوله‌ها و نرخ شکست در آن‌ها وجود داشته که مؤید این مطلب است که قطر لوله‌ها با نرخ شکست در لوله‌ها رابطه عکس دارد. در تحقیق حاضر، با تغییراتی که در فرم ابتدایی رابطه‌ی (۱) اعمال شد، عواملی مانند جنس لوله، طول لوله، فشار داخلی لوله و قطر لوله به معادله‌ی اولیه که فقط شامل زمان بود اضافه شدند. مدل اصلاح شده در رابطه‌ی (۳) آمده است.

$$N = \beta_0 + \beta_1(L) + \beta_2(P) + \beta_3(D) + \beta_4(Age) + \beta_5(AC) + \beta_6(DI) + \beta_7(GCI) + \beta_8(PE) + \beta_9(PVC) \quad (3)$$

که در آن N تعداد شکست‌ها، P فشار داخل لوله، D قطر لوله، Age سن لوله، AC آزیست، DI داکتایل، GCI چدن، PE پلی‌اتیلن و PVC لوله‌ی پی‌وی‌سی، است.

۲-۲- مدل رگرسیون نمایی

این مدل نوعی مدل رگرسیون غیرخطی است. کاربرد رگرسیون غیرخطی برای سطح وسیع تری از توابع است. فرم کلی یک مدل رگرسیون غیر خطی عبارت است از:

$$y = f(x, \beta) + \varepsilon \quad (4)$$

که y متغیر وابسته، $f(x, \beta)$ یک تابع غیرخطی با پارامترهای β_0 و β_1 و ... و ε میزان خطای باقی‌مانده است. این مدل معمولاً به مدلی که پارامترهای رگرسیون آن، خطی است، تغییر شکل می‌یابد و سپس مثل یک مدل رگرسیون خطی برازش داده می‌شود و فرضیات یک مدل رگرسیون خطی در مورد مدل تغییر شکل یافته اعمال می‌شود. بزرگ‌ترین مزیت مدل‌های غیرخطی این است که می‌توانند طیف وسیعی از توابع را برازش دهند. Shamir & Howard تحلیل رگرسیون غیرخطی را برای یافتن رابطه‌ی نمایی بین سن لوله و شکست‌های آن به کار بردند. مدل ارائه شده‌ی آن‌ها در معادله‌ی (۵) نشان داده شده است [7].

$$N(t) = N(t_0) \times e^{A(t+g)} \quad (5)$$

$N(t)$ تعداد شکست‌ها در واحد طول در سال، $N(t_0)$ تعداد شکست‌ها در واحد طول در سال نصب لوله، t زمان بین شکستگی یک سال تا سال شکستگی قبلی، g سن لوله در زمان t و A ضریب نرخ شکستگی در yr^{-1} است. این معادله با اضافه کردن دو پارامتر مؤثر دیگر به شکل معادله‌ی (۶) کامل شده است:

$$N(t) = C_1 \times C_2 \times N(t_0) \times e^{A(t+g)} \quad (6)$$

که C_1 بیانگر تأثیر شکست‌های قبلی لوله، بر اساس مشاهداتی است که مؤید این موضوع‌اند که لوله‌ای که قبلاً دچار حداقل یک شکست شده باشد، احتمال شکست مجدد آن بیشتر است و C_2 بیانگر تفاوت نرخ شکست در لوله‌های چدنی با قطر متفاوت است.

۲-۳- مدل خطی تعمیم یافته‌ی پواسون

مدل‌های خطی تعمیم یافته (GLM)، کاربرد رگرسیون خطی را برای اعمال روی داده‌های شمارشی، تعمیم می‌دهد. این مدل‌ها، پاسخ میانگین یک توزیع شرطی خاص را به یک تابع پیش‌بینی مرتبط می‌کنند. که مبتنی بر یک تابع توزیع احتمال (pdf) فرضی برای داده‌های گسسته (شمارشی) و همچنین یک تابع اتصال هستند که پارامترهای تابع توزیع احتمال را به متغیرهای موجود ارتباط می‌دهند [8]. مدل پواسون مدلی است که به طور متداول برای آنالیز رگرسیون داده‌های شمارشی مثل شکست در یک سیستم زیرساختی مورد استفاده قرار می‌گیرد. اگر $\bar{X}_i = [x_{1i}, x_{2i}, \dots, x_{ni}]$ بردار متغیرهای کمکی از قطعه‌ی i ام سیستم ($i = 1, 2, \dots, m$) باشد، تعداد شکست‌های قطعه‌ی i ام با y_i نمایش داده می‌شود. سیستم مورد نظر در این تحقیق، سیستم توزیع آب و قطعات آن قطعات لوله‌ها خواهند بود.

یک مدل رگرسیونی بر مبنای توزیع پواسون برای آماره‌ها، به شرط مقادیر مشاهده شده‌ی متغیرهای مستقل مشخص می‌کند که میانگین شرطی

¹ PROBABILITY DISTRIBUTION FUNCTION

آمارها در یک تابع پیوسته $\mu(\vec{x}_1, \vec{\beta})$ از مقادیر متغیرهای مستقل، طبق معادله‌ی شماره‌ی (۷) داده شده است، که در آن $\vec{\beta}$ بردار $n \times I$ از پارامترهای رگرسیون است.

$$E(y_i | x_i) = \mu(\vec{x}_1, \vec{\beta}) \quad (۷)$$

با داشتن \vec{x}_1 ، تابع چگالی احتمال مفروض برای y_i در مدل رگرسیون پواسون، برای مقادیر صحیح y_i ، به شکل معادله‌ی (۸) است.

$$\text{Logit}[P(x)] = 0.0009362L - 0.1993561P + 0.02521Age - 1.7572DI + 1.3791GCI \quad (۸)$$

تابع اتصال لگاریتمی در اینجا برای تعیین میانگین شرطی، برای مثال، به شکل زیر فرض شده است:

$$E(y_i | x_i) = \exp(\vec{x}_1, \vec{\beta}) \quad (۹)$$

در روش‌های رگرسیون خطی تعمیم یافته فرض بر آن است که متغیرهای توصیفی، مستقل‌اند. اما فرض توزیع نرمال باقی‌مانده‌ها در این روش وجود ندارد. هرچند که در مدل تعمیم یافته‌ی خطی پواسون فرض بر این است که میانگین شرطی و واریانس شرطی، برابرند:

$$\mu_i = \exp(\vec{x}_1, \vec{\beta}) \quad (۱۰)$$

مدل خطی تعمیم یافته‌ی پواسون در اینجا به شکل معادلات (۱۱) و (۱۲) و (۱۳) است:

$$P(Y = y | \vec{x}) = \frac{e^{-\mu} \times \mu^y}{y!} \quad (۱۱)$$

$$\mu = E(Y | \vec{x}) \quad (۱۲)$$

$$\text{Log}(\mu) = \beta_0 + \beta_1(D) + \beta_2(AC) + \beta_3(GCI) + \beta_4(DI) + \beta_5(PE) + \beta_6(PVC) + \beta_7(L) + \beta_8(Age) + \beta_9(P) \quad (۱۳)$$

که در آن Y تعداد شکست‌ها، β_i پارامترهای رگرسیون‌اند و متغیرهای مستقل، مطابق آنچه در بخش قبل توضیح داده شده، می‌باشند.

۲-۴- مدل خطی تعمیم یافته‌ی لجستیک (رگرسیون لجستیک)

این روش نوعی دیگر از مدل تعمیم یافته‌ی خطی است. این مدل احتمال خروجی‌های گسسته، مثل عضویت گروه، از یک دسته از متغیرهای توصیفی که امکان دارد گسسته، پیوسته، دویخی یا ترکیبی از این دو باشند را فراهم می‌سازد. در کل، متغیر پاسخ، دو بخشی است. مثلاً حضور یا غیاب، موفقیت یا شکست. در بسیاری از موارد دغدغه‌ی شرکت‌های تأمین‌کننده‌ی آب بررسی وقوع یا عدم وقوع شکست در یک لوله در یک بازه‌ی زمانی خاص است و نه تعداد شکست‌ها در آن لوله. گاهی حضور یک شکست دلیل نیاز به تعمیرات و صرف هزینه‌های زیاد برای آن است. متغیر وابسته یک متغیر باینری است که وقتی حداقل یک شکست در یک بازه‌ی زمانی، در لوله‌ای رخ دهد مقدار متغیر پاسخ ۱ خواهد بود. در این مدل لزومی ندارد متغیر مستقل توزیع نرمال داشته باشد. متغیرهای مستقل می‌توانند مقدار ۱ را با احتمال P و مقدار ۰ با احتمال $(1-P)$ داشته باشند. در مدل لجستیک ارتباط بین متغیرهای مستقل و وابسته خطی نیست. یک تابع رگرسیون لجستیک که تبدیل یافته‌ی P به لجیت^۱ آن است در معادله‌ی (۱۴) آمده است.

$$P = \frac{e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)}}{1 + e^{(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i)}} \quad (۱۴)$$

در اینجا α پارامتر ثابت رگرسیون، β_i ضرایب رگرسیون برای متغیرهای توصیفی و x_i ها متغیرهای مستقل‌اند. یک فرم جایگزین برای این مدل، مدل لجیت است که در آن تابع اتصال یک تابع لجیت است. این تابع در معادله‌ی (۱۵) نشان داده شده است.

$$\text{Logit}[P(x)] = \alpha + \beta_1(X_1) + \beta_2(X_2) + \dots + \beta_i(X_i) \quad (۱۵)$$

مدل لجستیک تعمیم یافته‌ی خطی که پیشنهاد شده در معادله‌ی (۱۶) آمده است:

$$\text{Logit}[P(x)] = \alpha + \beta_0(D) + \beta_1(AC) + \beta_2(GCI) + \beta_3(DI) + \beta_4(PE) + \beta_5(PVC) + \beta_6(L) + \beta_7(Age) + \beta_8(P) \quad (۱۶)$$

$P(x)$ احتمال وقوع شکست، $1-P(x)$ احتمال عدم وقوع شکست، α عرض از مبدأ و β_i ها پارامترهای رگرسیونی‌اند که برآورد می‌شوند.

^۱ LOGIT

۳. پایلوت مطالعاتی

پایلوت مورد مطالعه، ناحیه ی ۱ از نواحی سه گانه منطقه شمیرانات تهران است. این ناحیه در سال ۱۳۸۶ تعداد ۴۸۵۲۱ مشترک داشته است و تعداد حوادث ثبت شده از تیرماه ۱۳۸۳ تا آذرماه ۱۳۸۶ برابر با ۶۵۰۰۰ مورد می باشد [9]. بنابراین بالا بودن آمار حوادث در این ناحیه ضرورت امر تحقیق در این زمینه را نشان می دهد. به دلیل محدودیت های موجود در جمع آوری اطلاعات، در نهایت پارامترهای جنس، قطر، طول، فشار و سن لوله جهت بررسی در این تحقیق انتخاب شدند. در منطقه ی مورد مطالعه عمدتاً از لوله هایی با جنس های داکتایل، آزیست، چدن، پلی اتیلن و PVC استفاده شده است و میانگین سن لوله ها (تا آذرماه سال ۱۳۸۶) حدود ۲۹ سال می باشد. خلاصه ای از داده های ورودی در جدول (۱) آمده است.

جدول ۱ - خلاصه ی داده های ورودی

متغیر	توضیحات	واحد	میانگین	انحراف معیار
NB	تعداد شکست	ندارد	۰/۸۰۱۵	۱/۲۹۵
D	قطر لوله	میلی متر	۱۱۴/۱۱	۴۸/۳۲
AC	سیمان آزیست	ندارد	۰/۱۹۶	۰/۳۹۷
DI	داکتایل	ندارد	۰/۵۵۸	۰/۴۹۶
GCI	چدن	ندارد	۰/۸۰۳	۰/۲۷۶
PE	پلی اتیلن	ندارد	۰/۱۶۱	۰/۳۶۸
Age	سن لوله	سال	۲۹/۱۷۷	۹/۶
P	فشار	اتمسفر	۳/۸۶	۰/۶۸۱
L	طول لوله	متر	۶۸۶/۷۴	۵۵۰/۳۷

۴. نتایج

با تحلیل آمار موجود و به کمک روش های رگرسیونی، رابطه ی میان پارامترهای تأثیرگذار و نرخ وقوع حوادث به وقوع پیوسته، به دست آمده است. همان طور که قبلاً گفته شد، در این تحقیق چهار روش رگرسیونی برای مدل سازی تعداد و احتمال شکست در لوله ها به کار رفته است. برای تحلیل مدل های حاصل از هر یک از روش ها، دو رویکرد کلی مورد استفاده قرار گرفت. در ابتدا مدل به تمام داده های موجود برازش یافت. سپس، ابتدا در یک مرحله از ۷۰٪ داده ها به عنوان داده های مدل سازی و از ۳۰٪ باقی مانده، به عنوان داده های آزمون استفاده شد. از آنجا که در داده های موجود، برای هر لوله تنها یک ردیف داده وجود داشت که مقادیر مشاهداتی متغیرهای پیش بینی را در برابر تعداد شکست های مشاهده شده نشان می داد، اگر لوله های دارای تعداد شکست های صفر از مجموعه ی داده ها حذف می شدند، دیگر هیچ شکست صفری در مجموعه ی داده ها باقی نمی ماند و در نتیجه نتایج مدل ها احتمالاً تحت تأثیر قرار گرفته و دچار بیش برآورد می شد. ضمناً تعداد داده ها به حدود نصف کاهش می یافت. لذا ردیف های مربوط به شکست های صفر در مجموعه ی داده ها حفظ شدند. در این قسمت نتایج استفاده از هر چهار مدل توضیح داده می شود:

۴-۱- مدل خطی

ابتدا داده ها به مدل خطی برازش داده شدند و با استفاده از آزمون فرض معناداری، p-value برای متغیرهای مختلف محاسبه شد، متغیرهایی که سطح

معناداری پایینی داشتند (با p -value بزرگتر از ۰/۱) در برازش داده‌ها به مدل بعدی حذف شدند، به عبارتی ارتباط معناداری بین آن‌ها و تعداد شکست‌ها وجود ندارد. این چرخه تا زمانی ادامه می‌یابد که تمام متغیرهای موجود در مدل سطح معناداری بالایی داشته باشند (p -value کوچک‌تر از ۰/۰۰۱). در نهایت مناسب‌ترین مدل خطی، مدلی بود که در معادله‌ی (۱۷) نشان داده شده است.

$$NB = 1.024 + 0.00073L - 1.293DI \quad (17)$$

۴-۲- مدل نمایی

برآزش یک مدل رگرسیون غیرخطی، به مقادیر اولیه برای پارامترهای مدل نیاز دارد. یک انتخاب نامناسب منجر به دستیابی به یک مدل غیر بهینه می‌شود. مدل رگرسیون نمایی در معادله‌ی (۱۸) آمده است:

$$NB = \exp(0.000924L - 0.13P - 0.0049D - 0.79AC - 0.895DI + 0.905GCI + 0.86PE) \quad (18)$$

در این معادله NB تعداد شکست‌ها در ۳/۵ سال، L طول لوله، P فشار داخلی لوله، AC جنس لوله‌ی آزیست، DI لوله با جنس داکتیل، GCI لوله‌ی چدن و PE جنس لوله‌ی پلی‌اتیلن است. بر اساس پارامترهای برآورد شده در معادله ۱۸ برای متغیرهای موجود، با افزایش طول لوله تعداد شکست‌ها افزایش یافته و بین قطر لوله و تعداد شکست‌های آن رابطه‌ی عکس وجود دارد و استفاده از جنس داکتایل با کاهش معناداری تعداد حوادث همراه بوده است.

۴-۳- مدل خطی تعمیم یافته‌ی پواسون

در این مدل نیز مشابه مدل‌های خطی و نمایی، مدل‌سازی با ۹ متغیر ابتدایی آغاز شده و سپس در طی عملیاتی گام به گام، در هر مرحله متغیری که سطح معناداری پایینی داشت، حذف شده و برآزش مجدداً با متغیرهای باقیمانده انجام شد و در نهایت تعداد ۵ متغیر کمکی در مدل شماره ۴ باقی ماندند. سطح معناداری روی ۰/۰۵ تعریف شده است. مدل به دست آمده در معادله (۱۹) آمده است.

$$\text{Log}\mu = 0.7158 + 0.0008889L - 0.1442P - 0.003343D - 1.585DI \quad (19)$$

متغیرهای موجود در مدل (۱۹) عبارت‌اند از: طول لوله (L)، فشار داخلی لوله (P)، قطر لوله (D) و جنس لوله‌ی داکتایل. با تعریف سطح معناداری روی ۰/۰۵، متغیرهای فشار و جنس لوله‌ی داکتایل بالاترین سطح معناداری را از لحاظ آماری دارند. در این مدل تعداد شکست‌ها با افزایش طول لوله افزایش و با افزایش قطر، کاهش می‌یابند. هم‌چنین استفاده از لوله‌های داکتایل منجر به کاهش تعداد شکست‌ها می‌گردد.

۴-۴- مدل خطی تعمیم یافته‌ی لجستیک

مدل لجستیک به طور ویژه‌ای برای کار با داده‌های باینری ۰ و ۱، در مواقعی که با تعداد زیادی از صفرها روبرو هستیم، طراحی شده است. به ازای هر قطعه لوله، یک دسته از متغیرهای اندازه‌گیری شده، شامل متغیرهای پاسخ و متغیرهای مستقل وجود دارد. چنانچه یک لوله در طی بازه‌ی زمانی ثبت داده‌های جمع‌آوری شده (۸۶-۸۳)، حداقل یک شکست را تجربه کرده باشد، متغیر پاسخ مقدار ۱ خواهد بود. داده‌های موجود با مدل لجستیک برازش شدند و نتایج برازش نشان می‌دهد که متغیرهای زیر از لحاظ آماری، در سطح ۰/۰۵ معنادارند:

- طول لوله
- فشار لوله
- سن لوله
- متغیرهای جنس لوله‌ی داکتایل و چدن

در معادله‌ی (۲۰) مدل نهایی با پارامترهای برآورد شده نشان داده شده است.

$$\text{Logit}[P(x)] = 0.0009362L - 0.1993561P + 0.02521Age - 1.7572DI + 1.3791GCI \quad (20)$$

با افزایش طول و سن لوله، تعداد شکست‌های مدل افزایش می‌یابد، هم‌چنین استفاده از لوله‌هایی با جنس چدن داکتایل باعث کاهش تعداد شکست‌ها و برعکس، استفاده از لوله‌های چدنی باعث افزایش شکست‌ها می‌گردد. سطح معناداری این آزمون روی ۰/۰۵ تعریف شده است. مقادیر پارامترهای رگرسیون نشان می‌دهد که در لوله‌های با طول بیشتر، احتمال تجربه‌ی شکست بالاتر است. هم‌چنین، پارامترهای مربوط به جنس لوله‌ها، گویای این مطلب‌اند که در لوله‌های با جنس داکتایل، احتمال شکست کمتر و در لوله‌های جنس چدن خاکستری، احتمال شکست بالاتر است.

۵. تحلیل نتایج

در جدول ۲ آماره‌های نکویی برازش، برای مدل‌های رگرسیون خطی و رگرسیون نمایی و مدل پواسون برای رکورد ۳/۵ ساله نشان داده شده است. هم‌چنین در جداول ۳ و ۶ میانگین مربعات خطاها در سه مدل بررسی شده، به ترتیب برای آزمون فرض و خطا و صحت سنجی متقابل، آمده است.

جدول ۲- خلاصه آماره‌های نکویی برازش برای مدل‌های رگرسیون خطی و نمایی و پواسون

GCV	Deviance	AIC	Log likelihood	
1.42	1101.88	2488.83	-1240.41	مدل رگرسیون خطی
1.32	1016.7	2434.16	-1209.08	مدل رگرسیون نمایی
1.32	991.36	1811.5	-900.75	مدل رگرسیون تعمیم یافته‌ی خطی پواسون

با در نظر گرفتن مقادیر AIC، انحراف (Deviance) و GCV، مدلی که کمترین مقادیر را در این سه آماره دارد، مدل مناسب‌تر است. از سوی دیگر، با در نظر گرفتن لگاریتم احتمال (LL)، مدل با بیشترین مقدار این آماره مدل بهتری است. بر اساس نتایج موجود به نظر می‌رسد که کیفیت مدل خطی و نمایی در پیش‌بینی‌ها تقریباً مشابه است و مدل پواسون داده‌ها را به شکل بهتری برازش می‌دهد.

۶. نتیجه‌گیری

در تحقیق انجام شده از چهار مدل آماری مختلف تحت عنوان‌های مدل رگرسیون خطی، مدل رگرسیون نمایی، مدل رگرسیون خطی تعمیم یافته‌ی پواسون و مدل رگرسیون خطی تعمیم یافته‌ی لجستیک، برای بررسی شکست در لوله‌ها استفاده شده است. بررسی نتایج برازش داده‌ها با مدل خطی نشان می‌دهد که مدل خطی، برای مدل کردن قابلیت اعتماد سیستم شبکه‌ی توزیع آب مناسب نیست. مقدار R^2 در این مدل برابر ۰/۱۸ است که مقدار بسیار کمی بوده و این مسئله حاکی از غیرخطی بودن رابطه‌ی بین شکست‌ها و متغیرهای موجود می‌باشد. مدل نمایی ارائه شده در رابطه‌ی ۱۸، نشان می‌دهد که با افزایش طول لوله، تعداد شکست‌ها افزایش می‌یابد و با افزایش مقادیر متغیرهای قطر، تعداد شکست‌ها کاهش می‌یابد. هم‌چنین استفاده از لوله‌ی داکتایل باعث کاهش شکست‌ها و برعکس استفاده از لوله‌های با جنس سیمان آریست، چدن و پلی‌اتیلن، رابطه‌ی مستقیم با افزایش تعداد شکست‌ها دارد.

مقایسه‌ی معیارهای نکویی برازش در سه مدل خطی، نمایی و پواسون نشان می‌دهد که مقادیر AIC، انحراف (D) و لگاریتم درست‌نمایی برای مدل پواسون نسبت به دو مدل دیگر در وضعیت بهتری قرار دارد. هم‌چنین مقدار انحراف و AIC در این مدل کمتر از دو مدل دیگر است. در نهایت از مدل رگرسیون خطی تعمیم یافته‌ی لجستیک، استفاده گردید و محاسبات نشان داد که این مدل به دلیل اینکه کلیه متغیرهای موثر در شکست لوله‌ها در آن وجود دارند بصورت مناسب‌تری می‌تواند شکست لوله‌ها را پیش‌بینی نماید. در مجموع می‌توان گفت مدل لجستیک، مدل مناسب‌تری در تخمین و پیش‌بینی احتمال شکست لوله‌های سیستم توزیع پایلوت مورد نظر در طی بازه سه سال و نیم می‌باشد. در تمام مدل‌های ارائه شده، متغیر جنس لوله‌ی



شرکت مهندسی آب و فاضلاب کتور

کنگره علوم و مهندسی آب و فاضلاب ایران

دانشگاه تهران، تهران

۲۶ و ۲۷ بهمن ماه ۱۳۹۵



چدن داکتایل عامل قابل توجهی در کاهش تعداد شکست‌ها، در مقایسه با دیگر متغیرهای مستقل می‌باشد. برای به دست آوردن نتایج با قطعیت بیشتر نیاز به استفاده از داده‌های با بازه زمانی بیشتر می‌باشد که در این تحقیق امکان آن میسر نبود و ضرورت دارد که با به دست آوردن داده‌های واقعی در بازه‌های زمانی طولانی‌تر، تحقیق را ادامه داد. این نکته نیز لازم است گفته شود که در این تحقیق اطلاعات مربوط به فشار در پایلوت نیاز به تحقیق بیشتر داشته است که حتما در ادامه تحقیق می‌باید مد نظر قرار گیرد تا به نتایج دقیق‌تری در رابطه با ارتباط بین فشار و شکست لوله‌ها دست یافت.

۷. قدردانی

از شرکت آب و فاضلاب منطقه یک تهران برای همکاری و استفاده از داده‌های شکست لوله‌ها تشکر می‌شود.

۸. مراجع

۱. بیگی، ف. (۱۳۷۸). آسیب شناسی شبکه‌های توزیع آب شهری. مجله‌ی آب و محیط‌زیست، شماره ۳۷، ۱۰-۱۶.
۲. تشییعی، ح. (۱۳۸۴). استقرار و پیاده‌سازی نظام نگهداری و تعمیرات پیشگیرانه تأسیسات آب شرب کشور، چالش‌ها و دستاوردهای حاصله. سومین کنفرانس ملی نگهداری و تعمیرات.
3. Kettler, A., Goulter, I. (1985). "An analysis of pipe breakage in urban water distribution networks." Canadian Journal of Civil Engineering, 12(2), 286-293.
4. Shamir, U., Howard. C. (1979). "An analytical approach to scheduling pipe replacement." American Water Works Association, 248-258.
5. Andreou, S. A., Marks, D. H. & Clark, R. M. (1987). A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory. Advances in Water Resources, 10(1), 2-10 .
6. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). "Introduction to linear regression analysis." (Vol. 821): John Wiley & Sons.
7. Shamir, U., Howard. C. (1979). "An analytical approach to scheduling pipe replacement." American Water Works Association, 248-258.
8. Agresti, A., Kateri, M. (2011). "Categorical data analysis" , Springer.
- ۹- جلیلی قاضی زاده، م.، حنیفی یزدی، س. ح.، راستی اردکانی، ر. (۱۳۸۷). ارائه روابط پیش‌بینی وقوع حوادث در شبکه‌های توزیع آب شهری. دومین همایش ملی آب و فاضلاب با رویکرد بهره‌برداری.