طبیعی تأثیر می‌گذارد. (۲) محصولات تراریخته ممکن است در محیط زیست و تنوع زیستی تأثیراتی داشته باشند. علاوه بر این، جریان ترانس ژن‌ها ممکن است پیامدهای اقتصادی داشته باشد، در صورتی که محصول برداشت شده نتواند به‌عنوان محصول عاری از ژن تراریختگی فروخته شود. البته به چه میزان عواقب انتقال ژن رخ خواهد داد به صفات منتقل‌شده، دریافت‌کننده، و محیط زیست بستگی دارد. در حال حاضر انتقال ژن بین گونه‌ها به عواقبی مانند ایجاد علف‌های هرز جدید یا به خطر انداختن گونه‌های نادر تقریباً برای بسیاری از گیاهان زراعی اثبات‌شده است.

ادامه دارد ...

مهندس سجاد طلایی

کارشناس مجتمع تحقیقات کاربردی و تولید بذر

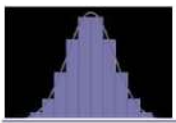
شرکت توسعه کشت دانه‌های روغنی

کاربرد مدل‌های آماری در اصلاح نباتات

کاربرد رگرسیون در اصلاح نباتات

به مدل‌هایی که با استفاده از توابع ریاضی، متغیر (های) مستقل، تغییرات متغیر وابسته را به‌طور کامل و دقیق بیان می‌کنند مدل قطعی (Deterministic Model) می‌نامند. مفروضات کلی مدل‌های رگرسیونی مشابه پیش فرض‌های تجزیه واریانس یعنی نرمال بودن باقی مانده‌ها، توزیع تصادفی خطاها، استقلال خطاها و ثابت بودن واریانس است.

رگرسیون ساده خطی توصیف‌کننده تغییرات یک متغیر وابسته بر پایه یک متغیر مستقل می‌باشد. مثلاً با استفاده از سطوح بیماری عملکرد را پیش‌بینی کرد. در اینجا سطوح بیماری مستقل در نظر گرفته‌شده است و تغییرات عملکرد نسبت به قطر ساقه گیاه بررسی می‌شود. توجه داشته باشید که اینجا تغییرات عملکرد نسبت به سطوح بیماری سنجیده می‌شود چون این بیماری است که روی عملکرد



تأثیر می‌گذارد ولی عملکرد روی سطوح بیماری تأثیری ندارد. پس سطوح بیماری مستقل و عملکرد وابسته در نظر گرفته می‌شود. ضریب رگرسیون فقط یک معیار برای ارتباط خطی است و هیچ نشانی از اینکه کدام خط راست بهترین است ارائه نمی‌دهد و فقط نشان‌دهنده بهترین ارتباط است. آنالیز رگرسیون ساده خطی به معنی تعیین این ارتباط است. لفظ ساده اشاره به این حقیقت دارد که یک متغیر مستقل در مدل وجود دارد. و ربطی به پیچیده یا ساده بودن آنالیز ندارد.

در رگرسیون خطی ساده سعی در ایجاد رابطه علت، معلولی خطی بین متغیر مستقل و متغیر وابسته می‌باشد ولی در بسیاری از موارد ضرورت دارد که از چند متغیر مستقل استفاده شود. در این موارد از رگرسیون چندگانه استفاده می‌شود. رگرسیون خطی چندگانه به دنبال یافتن تابع یا مدلی می‌باشد که تغییرپذیری متغیر وابسته را به‌وسیله بیش از یک متغیر مستقل به‌خوبی بیان کند. از رگرسیون خطی چندگانه می‌توان به انواع رگرسیون پیش‌رو، پس‌رو، گام‌به‌گام و غیره اشاره کرد. در این رابطه باید به هم‌راستایی خطی و داده‌های پرت نیز توجه شود. هم‌راستایی خطی چندگانه در شرایطی ایجاد می‌شود که همبستگی بین متغیرهای مستقل وجود داشته باشد. در این شرایط برآوردها اعتباری ندارند زیرا واریانس برآوردها بزرگ است. هم‌راستایی خطی را می‌توان با آماره عامل تورم واریانس (VIF) تشخیص داد. برای تصحیح هم‌راستایی خطی چندگانه استفاده از راه‌حل‌های زیر پیشنهاد می‌گردد:

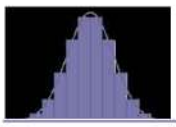
متغیرهای مستقل غیرضروری از مدل حذف شوند.

چند متغیر مستقل وابسته را به‌عنوان یک متغیر مستقل جدید تعریف نمود.

مشاهده‌های مشکل‌ساز را حذف نمود.

از رگرسیون رایج و تجزیه به مؤلفه‌های اصلی استفاده نمود.

در برخی شرایط اثر یک متغیر مستقل بر متغیر وابسته، خطی نیست. در این موارد پیش‌بینی حاصل از مدل صحیح نمی‌باشد و باید رگرسیون درجه دو نیز بررسی گردد. برای این کار می‌توان به رگرسیون چندجمله‌ای، غیرخطی و تکه‌ای (Segmented) استفاده کرد. در رگرسیون چندجمله‌ای از توابع درجه ۲، ۳ و بالاتر می‌توان استفاده نمود. گاهی رابطه بین متغیر مستقل و وابسته یک رابطه غیرخطی حقیقی است. مدل‌های رگرسیون غیرخطی به‌وسیله پارامترهای غیرخطی معین می‌شوند و می‌توانند شکل‌های مختلفی داشته باشند که ترکیبی از پارامترهای خطی و غیرخطی هستند. توابع نمایی، لگاریتمی و لجستیک مثال‌های از توابع غیرخطی است که برای برازش پدیده‌های بیولوژیکی مورداستفاده قرار می‌گیرند.



رگرسیون تکه‌ای را می‌توان برای برآورد نیازمندی‌های تغذیه‌ای استفاده کرد. از یک نقطه مشخص، مقادیر متغیر وابسته ثابت می‌شود. یعنی با تغییر متغیر مستقل مقادیر وابسته دیگر تغییری نمی‌کنند. مثلاً افزایش ویتامین تا یک نقطه خاصی باعث افزایش وزن می‌گردد و هدف ما به دست آوردن این نقطه یا گره می‌باشد.

برآورد پارامترهای رگرسیونی با دو روش برآورد حداقل مربعات و برآورد بیشینه درست نمایی (Maximum Likelihood Estimation) برآورد نمود. جهت آزمون فرض‌های مرتبط با پارامترها یعنی پارامتر شیب‌خط از آزمون T –student و آزمون نسبت درست نمایی (LR) استفاده می‌شود. از آماره ضریب تبیین، ضریب تبیین تصحیح‌شده (R2adj)، مالو (Cp)، خطای جذر میانگین مربعات (RMSE)، معیار اطلاعات آکائیک (AIC)، معیار اطلاعات بیزین، شوارتس (BIC) می‌توان برای ارزیابی مدل رگرسیونی مورداستفاده قرار می‌گیرند. ضریب تبیین معیاری برای اندازه‌گیری کفایت مدل رگرسیون است. برخی از پژوهشگران استفاده از ضریب تبیین تصحیح‌شده را ترجیح می‌دهند. از آماره مالو برای قضاوت درباره یک معادله به‌جای یک میانگین مربعات انحراف از مدل، میانگین مربعات خطای مقدار پیش‌بینی‌شده در نظر گرفته می‌شود. آماره خطای جذر میانگین مربعات تفاوت بین مقدار پیش‌بینی‌شده توسط مدل آماری و مقدار واقعی می‌باشد و یک ابزار مناسب برای مقایسه خطاهای پیش‌بینی است و برای مقایسه چند مجموعه داده کاربرد ندارد. معیار اطلاعات آکائیک و بیزین- شوارتس تعادلی میان دقت مدل و پیچیدگی آن برقرار می‌کنند. معیار اطلاعات آکائیک یک‌میزان از کیفیت نسبی مدل آماری از یک مجموعه از داده‌ها می‌باشد. درواقع معیار اطلاعات آکائیک ابزاری برای انتخاب مدل است. این معیار یک معادله بین برازش و پیچیدگی مدل را توضیح می‌دهد. این آماره بر اساس پراکنش اطلاعات بناشده است.کمتر بودن این مدل‌ها در یک مدل نسبت به مدل دیگر گویای آن است که افزودن یک متغیر جدید هزینه ناشی از کاهش کارایی (به دلیل افزایش تعداد متغیرها) را از طریق کاهش به‌اندازه کافی SSR جبران نموده است. برای ۱۰۰ <N، ضابطه BIC شاخص سخت‌گیرانه‌تری نسبت به AIC خواهد بود. (Wooldridge) مدل‌هایی که به‌صورت نوعی دارای مقدار آماره آکائیک، بیزین، خطای جذر میانگین مربعات و مالو کوچک‌تر و دارای ضریب تبیین، ضریب تبیین تصحیح‌شده بالاتری دارند ترجیح داده می‌شوند.