آمار و احتمال با پایتون(۲)

نجمه مدنی

آزمون فرضیه

- برای بررسی صحت یک فرضیه از آزمون فرضیه استفاده می کنیم.
- اولین قدم تعیین فرضیه صفر و فرضیه مقابل است.
- معولا فرضی که انتظار امید می رود که رد شود به عنوان فرضیه صفر مطرح می شود.
- فرضیه صفر و فرضیه مقابل نقطه اشتراکی با یکدیگر ندارند.
- هدف از آزمون فرضیه این نیست که تعیین کنیم که فرضیه صفر یا فرضیه مقابل درست است.
- در آزمون فرضیه مشخص کنیم آیا شواهد موجود از رد فرضیه صفر (فرضیه مقابل؟) پشتیبانی می کنند یا خیر.

آزمون فرضیه

- **نمونه مسایلی** که از **آزمون فرضیه** برای یافتن پاسخ از آن ها استفاده می کنیم:
- تعیین پارامترهای یک مدل آماری (توزیع)
- آیا ذراتی که از یک ماده رادیواکتیو خارج می شود توزیع پواسن دارد؟
- آیا دو واقعه از هم مستقل هستند؟
- آیا مقدار یک پارامتر دریک تحقیق و بررسی علمی مساوی یک مقدار خاص است؟
- آیا IQ بستگی به تحصیلات والدین دارد؟
- آیا سیگار کشیدن طول عمر را کاهش می دهد؟

آزمون فرضیه

- آیا ممکن است در آزمون فرضیه دچار خطا شویم؟
- بله، با دو نوع خطا مواجه هستیم
- خطای نوع ۱: فرضیه صفر را رد کنیم در حالیکه فرضیه صفر درست است.
- خطای نوع ۲: فرضیه صفر را قبول کنیم در حالیکه فرضیه صفر (مقابل) نادرست (درست) است.

تعیین فرضیه صفر و فرضیه مقابل

- مثال: اداره راهنمایی و رانندگی در یک شهر بزرگ ادعا کرده است که ۸۰٪ از شرکت کنندگان در امتحانات رانندگی قبول می شوند، اما یک روزنامه با نظرسنجی از ۹۰٪ نوجوان شهر که به طور تصادفی انتخاب شده اند و در این آزمایش شرکت کرده بودند دریافت فقط ۶۱ نفر (۶۸٪) در امتحانات قبول شده اند. آیا این داده ها نشان می دهد که نرخ قبولی نوجوانان پایین تر از نرخ گزارش شده توسط راهنمایی و رانندگی است.
- فرض صفر: نرخ قبولی نوجوانان در امتحانات راهنمایی رانندگی ۸۰٪ است
- فرض مقابل: نرخ قبولی نوجوانان در امتحانات راهنمایی رانندگی کمتر از ۸۰٪ است.

تعیین فرضیه صفر و فرضیه مقابل

- شرکت خدماتی آ ادعا می کند میانگین مدت زمان انتظار مشتریان تا دریافت سرویس نیم ساعت یا کمتر است. شما با این ادعا موافق نیستید و به این منظور آزمونی انجام می دهید.
- فرضیه صفر: میانگین مدت زمان انتظار مشتریان شرکت آ برای دریافت سرویس مساوی یا کمر از نیم ساعت است.
- فرضیه مقابل: میانگین مدت زمان انتظار مشتریان شرکت آ برای دریافت سرویس مساوی بیش از نیم ساعت است.

تعیین فرضیه صفر و فرضیه مقابل

- تصور می شود که آب و هوای گرم از شیوع بیماری X می کاهد.
- فرضیه صفر: شرایط آب و هوایی در شیوع بیماری X تاثیر دارد.
- فرضیه مقابل: شرایط آب و هوایی در شیوع بیماری X تاثیری ندارد.

آزمون فرضیه:

- فرموله کردن آزمون فرضیه:
- متغیر تصادفی x دارای تابع توزیع $F_X(x,\theta)$ است که به پارامتر θ بستگی دارد. همچنین بردار نمونهn بعدی X از این توزیع موجود است.
- فرضیه صفر: $\theta = \theta_0$
- فرضیه مقابل: $\theta \neq \theta_0$
- تحت فرضیه صفر تابع چگالی $f_X(X,\theta_0)$ خواهد بود. قسمتی از فضای نمونه که مقدار تابع چگالی در آن ناچیز باشد را ناحیه بحرانی D_c و قسمت مکمل این ناحیه $\overline{D_c}$ را ناحیه قبول فرضیه صفر می نامیم.
- خطای نوع ۱: $\alpha = p\{\mathrm{X} \in D_c | H_0\}$
- خطای نوع ۲: $\beta = p\{\mathrm{X} \in \overline{D_c} | H_1\}$

آزمون فرضیه:

- معمولا برای یک سطح برای α مشخص می کنیم و سپس ناحیه بحرانی را بگونه ای تعیین می کنیم که خطای نوع دوم βکمینه شود. اگر مقدارβ باز هم بزرگ باشد مقدار α را به سطحی قابل تحمل افزایش می دهیم و اگر باز هم بزرگ بود n را افزایش می دهیم.
- جستجو در فضای نمونه (n بعدی) برای تعیین ناحیه بحرانی آسان نیست بنابراین از روش ساده تری استفاده می کنیم.

آزمون فرضیه:

- تابعی مانند $q = g(X)$ را انتخاب می کنیم برای تعیین این تابع که به آن آماره آزمون گفته می شود از تخمین گر های نقطه ای مانند میانگین نمونه یا واریانس نمونه استفاده می شود.
- تابع چگالی احتمال آماره آزمون تحت فرضیه صفر مشخص می کنیم.
- مانند آنچه توضیح داده شد مقداری برای α در نظر می گیریم و ناحیه بحرانی را بر اساس تابع چگالی q مشخص می کنیم.
- با استفاده از داده های موجود (نمونه جامعه/توزیع) مقدار آماره آزمون را مشخص می کنیم. چنانچه این مقدار در ناحیه بحرانی قرار داشت فرضیه صفر را رد می کنیم و در غیر این صورت آن را قبول می کنیم.

آزمون فرضیه:مثال

- می خواهیم این فرضیه را که میانگین متغیر تصادفی X برابر مقدار ... $H_0: \eta = \eta_0$ را امتحان کنیم. فرض می کنیم واریانس این متغیر تصادفی مشخص است. آماره آزمون بر اساس میانگین نمونه به صورت زیر تعریف می شود:

$$q = \frac{\bar{X} - \eta_0}{\frac{\sigma}{\sqrt{n}}}$$

- برای n های به اندازه کافی بزرگ $\bar{X} \sim N\left(\eta, \frac{\sigma^2}{n}\right)$ $q \sim N(\eta_q, 1)$
- تحت فرضیه H_0 $q \sim N(0,1)$

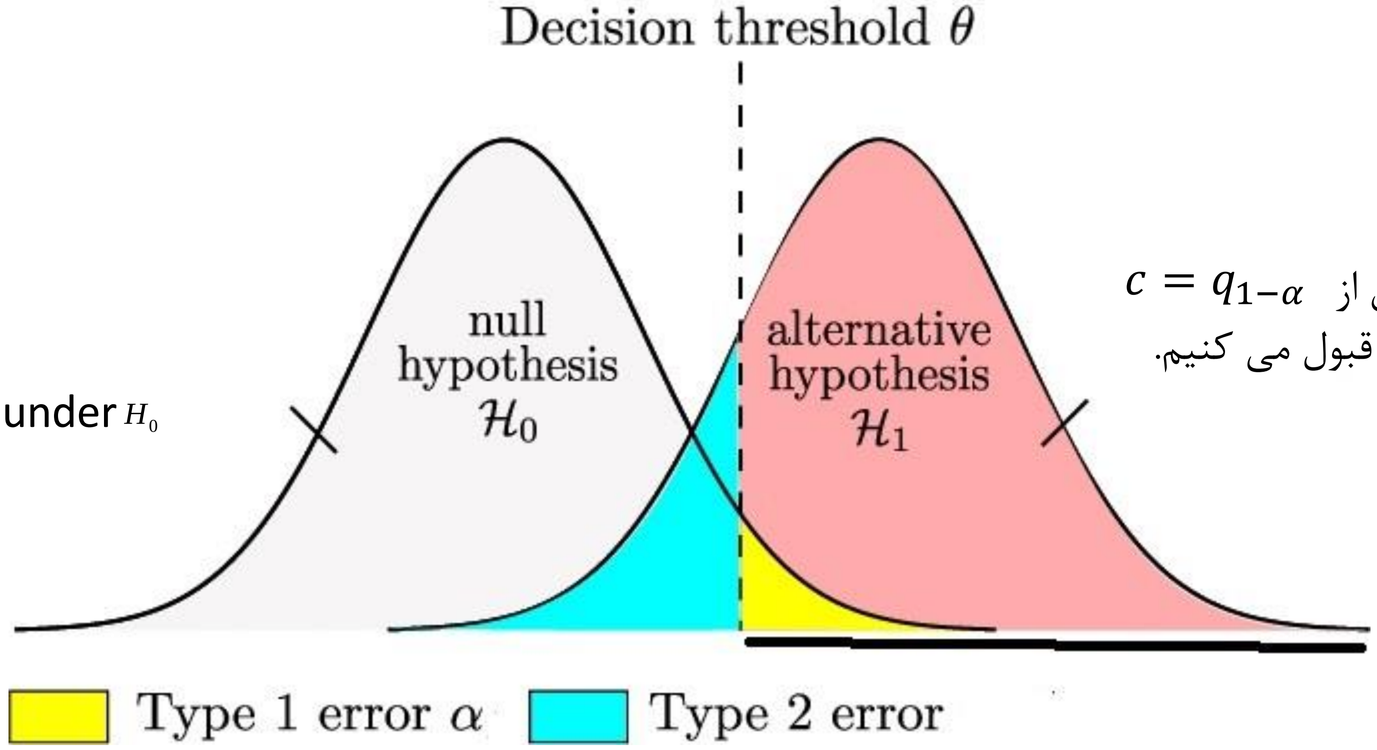
آزمون فرضیه:مثال

$H_0: \theta = \theta_0$
$H_1: \theta > \theta_0$

$\alpha = P\{q > c \mid H_0\}$

$c = q_{1-\alpha}$

q_u: is u percentile of q under H_0



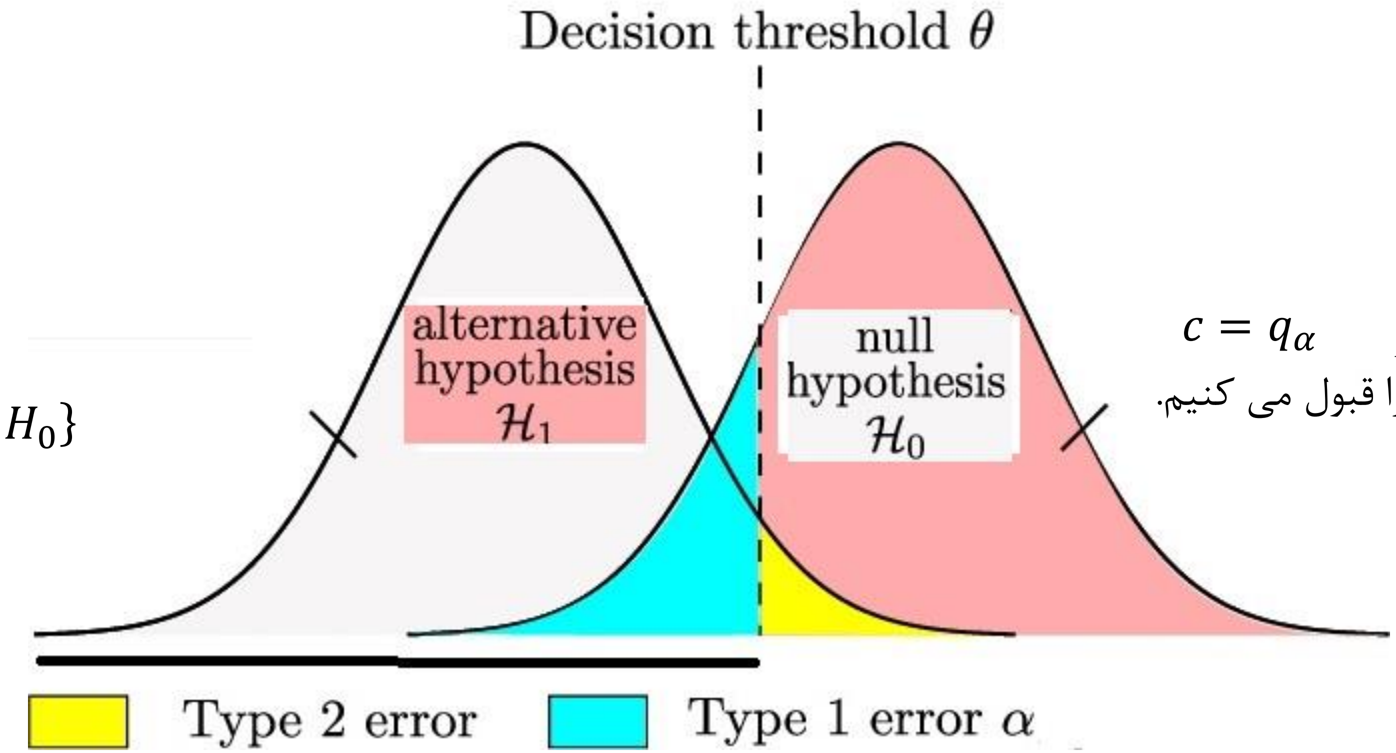
چنانچه مقدار آماره آزمون از $c = q_{1-\alpha}$
کمتر باشد فرضیه H_0 را قبول می کنیم.

Nhan, Le AU - Retraint, Florent
PY - 2018/12/01 SP - 331 EP - 335 T1 - Statistical Detector of Resampled TIFF ImagesDO - 10.1109/ISSPIT.2018.8642651

آزمون فرضیه:مثال

$H_0: \theta = \theta_0$
$H_1: \theta < \theta_0$

$\alpha = P\{q < c | H_0\}$

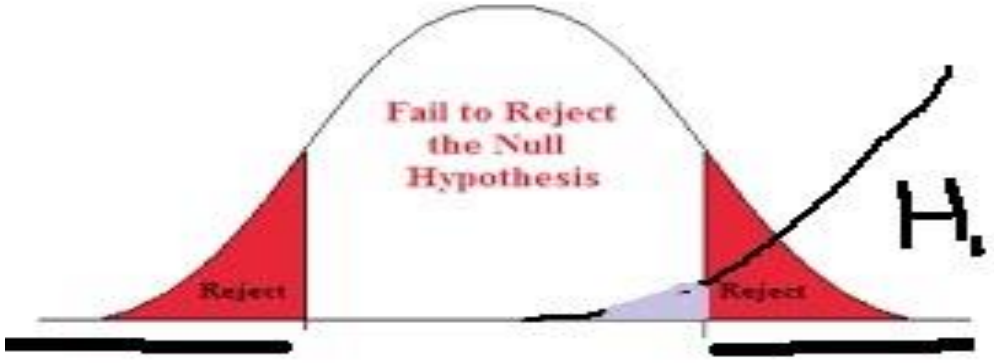


چنانچه مقدار آماره آزمون از $c = q_\alpha$ بیشتر باشد فرضیه H_0 را قبول می کنیم.

آزمون فرضیه:مثال

$H_0: \theta = \theta_0$
$H_1: \theta \neq \theta_0$

$$P\{q < c|H_0\} + P\{q > c|H_0\} = \alpha$$



فرضیه H_0 را قبول کنیم اگر و تنها اگر

$$q_{\frac{\alpha}{2}} < q < q_{1-\frac{\alpha}{2}}$$

آزمون فرضیه: مثال

- مثال: ولتاز منبع ولتاژی را ۲۵ بار اندازه می گیریم و نتیجه $\bar{X}= 110.12$
- می خواهیم فرضیه $v = v_0 = 110$ را در برابر $v = v_0 \neq 110$ امتحان کنیم. فرض می کنیم خطای اندازه گیری $N(0,.4^2)$ و $\alpha = 0.05$
- در این مسئله $z_{1-\frac{\alpha}{2}} = z_{0.975} = 2$

$$q = \frac{110.12 - 110}{\frac{.4}{\sqrt{25}}} = 1.5$$

- با توجه به اینکه $1.5 \in (-2,2)$ فرضیه صفر را قبول می کنیم.

آزمون فرضیه: مثال

- در صورتی که واریانس داده ها (درایه های بردار نمونه) مشخص نباشد از آماره آزمون زیر استفاده می کنیم:

$$q = \frac{\bar{X} - \eta_0}{\frac{S}{\sqrt{n}}}$$

- s^2 واریانس نمونه متغیر تصادفی x است. در این صورت و با فرض نرمال بودن x متغیر تصادفی q
- تحت فرضیه صفر داری توزیع t student با درجه n-1 است.
- در مثال قبل فرض می کنیم که واریانس مشخص نیست و با استفاده از اندازه گیری $s = 0.6v$

$$t_{1-\frac{\alpha}{2}}(n-1) = t_{0.975}(25) = 2.06 = -t_{0.025}$$

$$q = \frac{110.12 - 110}{\frac{0.6}{\sqrt{25}}} = 1$$

- با توجه به اینکه $1 \in (-2.06\ 2.06)$ فرضیه صفر را قبول می کنیم.

سوال

- بردار نمونه مناسب چگونه به دست می آید؟
- در حالی که می توانیم از تخمین گر ها برای تخمین پارمترهای توزیع استفاده کنیم لزوم استفاده از آزمون فرضیه در این مسائل چیست؟