



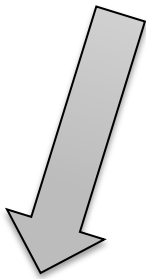
معرفی چند آزمون برای وجود روند در داده های تابعی وابسته به زمان

زهرة شیشه بر، سید مرتضی نجیبی، سکینه رمضانی
گروه آمار دانشگاه شیراز

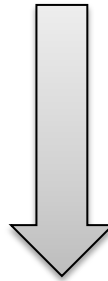
- ✓ بسیاری از پدیده‌ها ذاتاً توابع پیوسته‌ای هستند که مدل‌سازی آن‌ها به کمک روش‌های چند متغیره کلاسیک به خوبی انجام نمی‌گیرد.
- ✓ از این رو به جای تحلیل آن‌ها در فضای برداری با بعد متناهی، لازم است آن‌ها را در فضای تابعی با بعد نامتناهی در نظر بگیریم.
- ✓ چنین اقدامی به رویکرد جدیدی در آمار تحت عنوان تجزیه و تحلیل داده‌های تابعی *FDA* منتهی می‌شود.
- ✓ در این رویکرد داده‌ها به فضای تابعی با بعد نامتناهی متعلق هستند و لذا برای تحلیل آن‌ها ناگزیر به کاهش بعد هستیم.
- ✓ به منظور استفاده از *FDA* ابتدا لازم است مشاهدات را به صورت تابع پیوسته‌ای در نظر بگیریم.
- ✓ برازش منحنی‌های مناسب به مشاهدات با استفاده از روش‌های رگرسیون ناپارامتری مثل اسپلاین‌های هموارساز امکان‌پذیر است.

سری‌های زمانی تابعی (FTS)

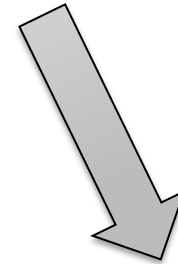
در تجزیه و تحلیل یک *FTS* سه موضوع مدنظر قرار می‌گیرد



مؤلفه‌ی تصادفی



تشخیص مؤلفه‌ی دوره-
های منظم



تشخیص مؤلفه روند

T تابع حقیقی را در نظر بگیرید که هر کدام از توابع در N نقطه به صورت زیر مشاهده شده اند:

$$(x_t(s_1), x_t(s_2), \dots, x_t(s_N)) \quad t = 1, \dots, T$$

مشاهدات رو بصورت توابع پیوسته به صورت زیر در نظر میگیریم

$$x_t(s_j) = \mu_t(s_j) + e_t(s_j), t = 1, \dots, T, j = 1, \dots, N$$

$$x_t(s) = \mu_t(s) + e_t(s), t = 1, \dots, T$$

روند افزایشی قوی برای FTS

$$\chi = \{x_t(s), s \in [\cdot, 1]\}_{t=1}^T$$

گوییم دنباله ی $\{x_t(s)\}_{t=1}^T$ دارای روند افزایشی است اگر:

$$x_1(s) \leq x_2(s) \leq \dots \leq x_t(s), s \in [\cdot, 1].$$

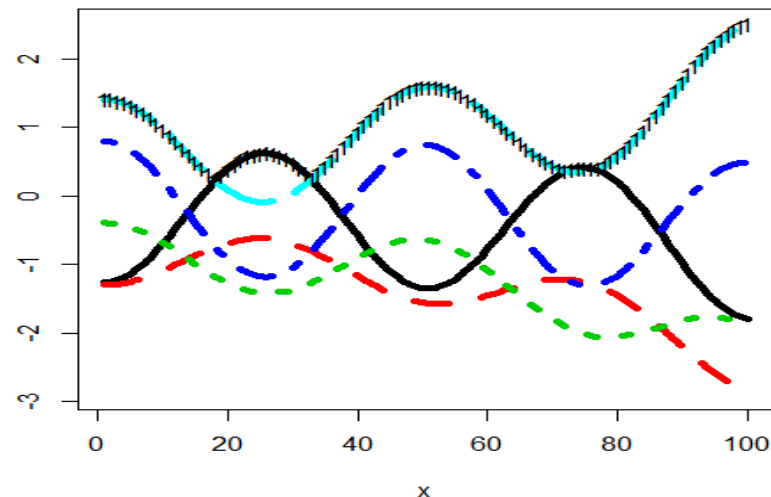
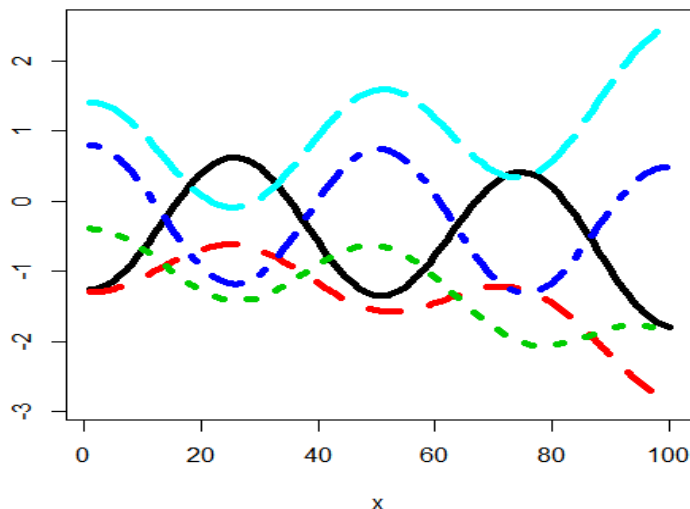
همانطور که می بینید انتظار مشاهده چنین یکنواختی در هر نقطه از برای تعیین روند در کل s فاصله ی زمانی، در عمل بسیار دور از واقعیت است.

تابع رکورد ماکسیمم

$$r_t : C[\cdot, 1]^t \rightarrow C[\cdot, 1],$$

$$r_t(s) \equiv \max(x_1(s), x_2(s), \dots, x_t(s)), \quad s \in [\cdot, 1]$$

✓ ایده این است که هر یک از منحنی‌های بعدی در دنباله، رکوردهایی که توسط منحنی‌های قبلی به دست آمده را ادامه دهند یا رکورد و سابقه جدیدی را ثبت کنند که نشان‌دهنده تدریجی روند افزایشی در دنباله است.



ایده فوق را به وسیله معیار ω_t توضیح می دهیم:

$$\omega_t \equiv \int \mathbf{I}_{\{r_t(s)=x_t(s)\}} ds.$$

در ادامه با استفاده از سه تعریف متفاوت روند قوی تا ضعیف، نحوه انجام آزمون زیر را پایه ریزی می کنیم:

$$\begin{cases} H_0 : & \text{"در یک سری از داده های تابعی روندافزایی وجود ندارد"} \\ H_1 : & \text{"otherwise"} \end{cases} \quad (1)$$

$$0 < \omega_2 - \frac{1}{2} < \omega_3 - \frac{1}{3} < \dots < \omega_T - \frac{1}{T}$$

تعریف ۱: روند افزایشی قوی:

$$H_b : \min_{t \in \square} (\omega_t - \frac{1}{t}) > 0, T \in \{2, \dots, T\}$$

تعریف ۲: روند افزایشی ضعیف:

$$\varpi_{[k]} > 0, \forall k \geq k_0$$

روند افزایشی k_0 ضعیف:

$$\varpi_{[2]} < \varpi_{[3]} < \dots < \varpi_{[T]}, \quad \varpi_t \equiv \omega_t - \frac{1}{t}, t = 1, \dots, T$$

آزمون‌های ناپارامتری برای روند افزایشی در FTS

متغیر تصادفی $\leftarrow \{X_t(s), s \in [\cdot, 1]\}_{t=1}^T$

$$R_t(s) = \max(X_1(s), X_2(s), \dots, X_t(s)) \text{ , } \cdot \leq s \leq 1 \text{ , } t \in \square \equiv \{2, \dots, T\}.$$

نسبت زمانی که تابع X_t برابر $R_t(s)$

$$W_t \equiv \int_{\cdot}^1 \mathbf{I}_{\{R_t(s) = X_t(s)\}} ds.$$

و مقدار مورد انتظار

$$J_t \equiv E(W_t) = \int_{\cdot}^1 P(X_t(s) = R_t(s)) ds.$$

انتظار داریم که اگر هیچ روندی وجود نداشته باشد یعنی به ازاء هر سال t داریم : $J_t = \frac{1}{t}$

با تعریف $\bar{J} \equiv (\bar{J}_2, \bar{J}_3, \dots, \bar{J}_T)$ و $\bar{J}_t = (J_t - \frac{1}{t})$ می توان آماره های ترتیبی آن را به صورت

$$\bar{J}_{[2]} < \bar{J}_{[3]} < \dots < \bar{J}_{[T]}.$$

بدیهی است که آزمون فرض $k_{\bullet}$ ضعیف به صورت زیر تبدیل می شود:

$$\begin{cases} H_{\bullet} : \bar{J}_{[k]} = 0, & \forall k \in \square \\ H_{A_{\bullet}} : \bar{J}_{[k_{\bullet}]} > 0, & \text{for some } k_{\bullet}. \end{cases} \quad (2)$$

دقت کنید که هر گاه $\bar{J}_{[k_{\bullet}]} > 0$ برای یک $k_{\bullet}$ انگاه $\bar{J}_{[k]} > 0, \forall k > k_{\bullet}$

$$\begin{aligned} &(x_{\backslash}(s_{\backslash}), x_{\backslash}(s_{\backslash}), \dots, x_{\backslash}(s_N)) \\ &(x_{\backslash}(s_{\backslash}), x_{\backslash}(s_{\backslash}), \dots, x_{\backslash}(s_N)) \\ &\vdots \\ &(x_T(s_{\backslash}), x_T(s_{\backslash}), \dots, x_T(s_N)) \end{aligned}$$

$$X_t(s_j) = \mu_t(s_j) + e_t(s_j), \quad t = \backslash, \dots, T, \quad j = \backslash, \dots, N$$

$$\tau_t \equiv \sum_{j=\backslash}^N \mathbf{I}_{\{R_t(s_j)=X_t(s_j)\}}, \quad \longrightarrow \tau_t \sim B(N, p)$$

$$\hat{J}_t \equiv \frac{\tau_t}{N} \qquad \longrightarrow \qquad \hat{J}_t - \frac{1}{t} > c_\alpha$$

$$\overline{\hat{J}} \equiv (\hat{J}_1 - \frac{1}{1}, \hat{J}_2 - \frac{1}{2}, \dots, \hat{J}_T - \frac{1}{T})$$

$$\overline{\hat{J}}_{[k]} > \cdot$$

محاسبه آماره آزمون و مقادیر بحرانی

با توجه به عدم استقلال متغیرهای تصادفی $\{(\hat{J}_t - 1/t), t = 1, \dots, T\}$ ، محاسبه مقادیر دقیق بحرانی به دلیل پیچیدگی‌های زیادی که در ساختمان ذات ترکیبیاتی مسئله وجود دارد، دشوار است.

جهت استفاده از آماره‌ی آزمون در انجام آزمون فرض‌های معرفی شده، ابتدا توزیع هر یک را با استفاده از شبیه‌سازی از مقدار آماره آزمون تحت فرض صفر تقریب می‌زنیم.

بر اساس جایگشت بین توابع و نقاط آنها با رویکردی که مشابهت زیادی با آزمون ناپارامتری جایگشت دارد، توزیع آماره آزمون فرض‌های معرفی شده را تقریب می‌زنیم.

گام
اول

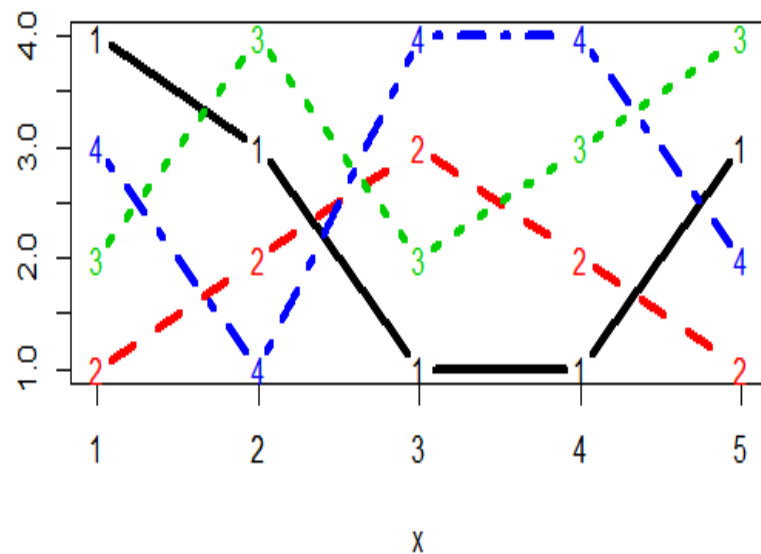
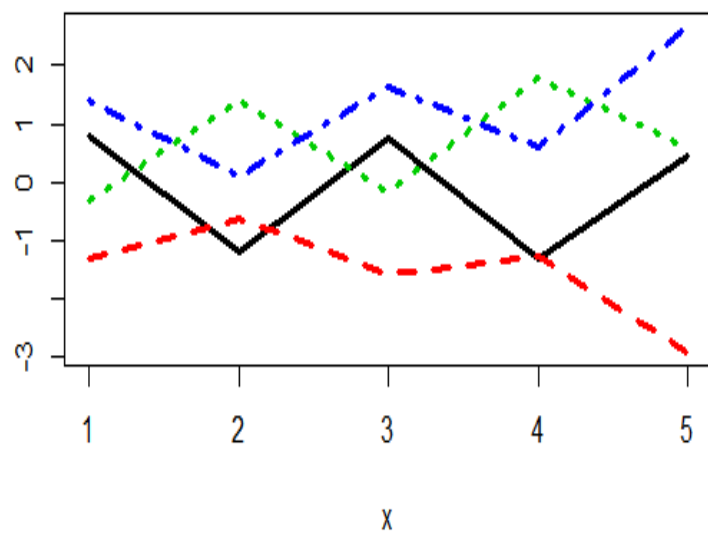
- تبدیل هر تابع به بردار رتبه متناظر خود

گام
دوم

- تصویر تابعی با جایگشت رتبه‌ها روی شبکه مشبک

گام
سوم

- محاسبه مقادیر آماره آزمون بر اساس بردار رتبه تحقق یافته تحت فرض صفر



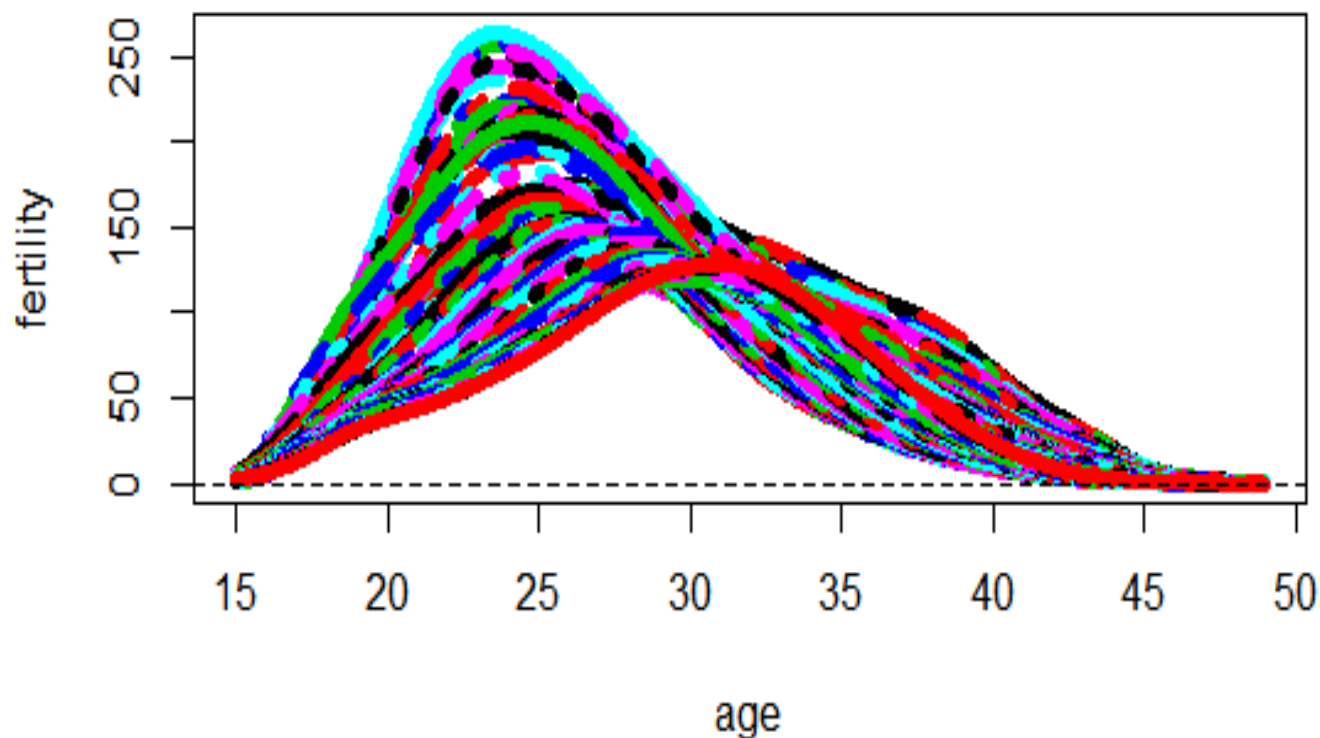
	$f_1(s_j)$	$f_2(s_j)$	$f_3(s_j)$	$f_4(s_j)$
s_1	4	1	2	3
s_2	3	2	4	1
s_3	1	3	2	4
s_4	1	2	3	4
s_5	3	1	4	2

بعد از محاسبه مقادیر $\hat{J}_t - 1/t, t = 1, \dots, T$ برای تمام منحنی‌ها می‌توان مقادیر $\hat{J}[k]$ را بر اساس سه گام گفته شده و با استفاده از بردار رتبه‌های تصادفی تولید شده، بدست آورد.

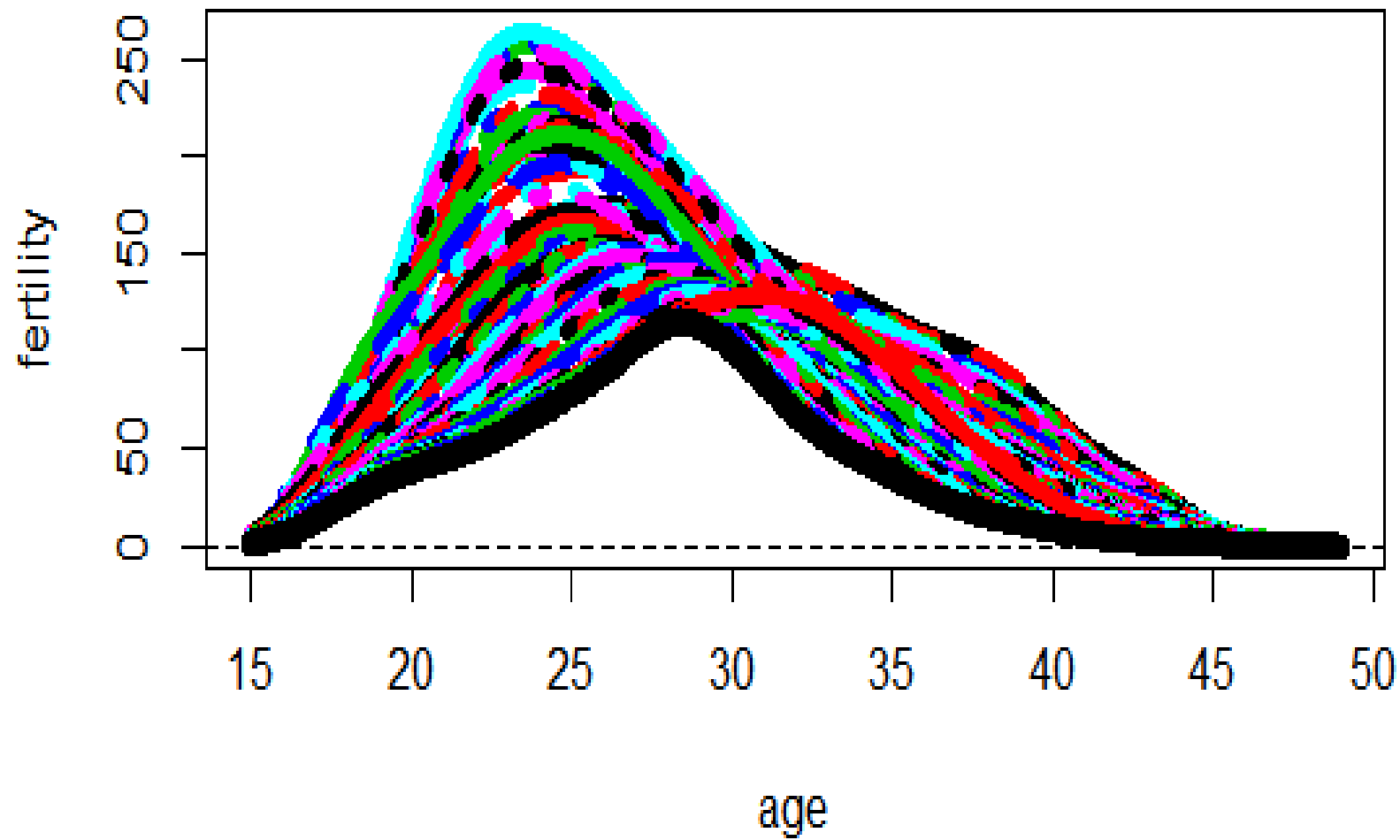
با n بار تکرار سه گام گفته شده و تولید رتبه‌های تصادفی در هر مرحله، تعداد n آماره آزمون تحت فرض صفر به صورت تصادفی تولید شده است.

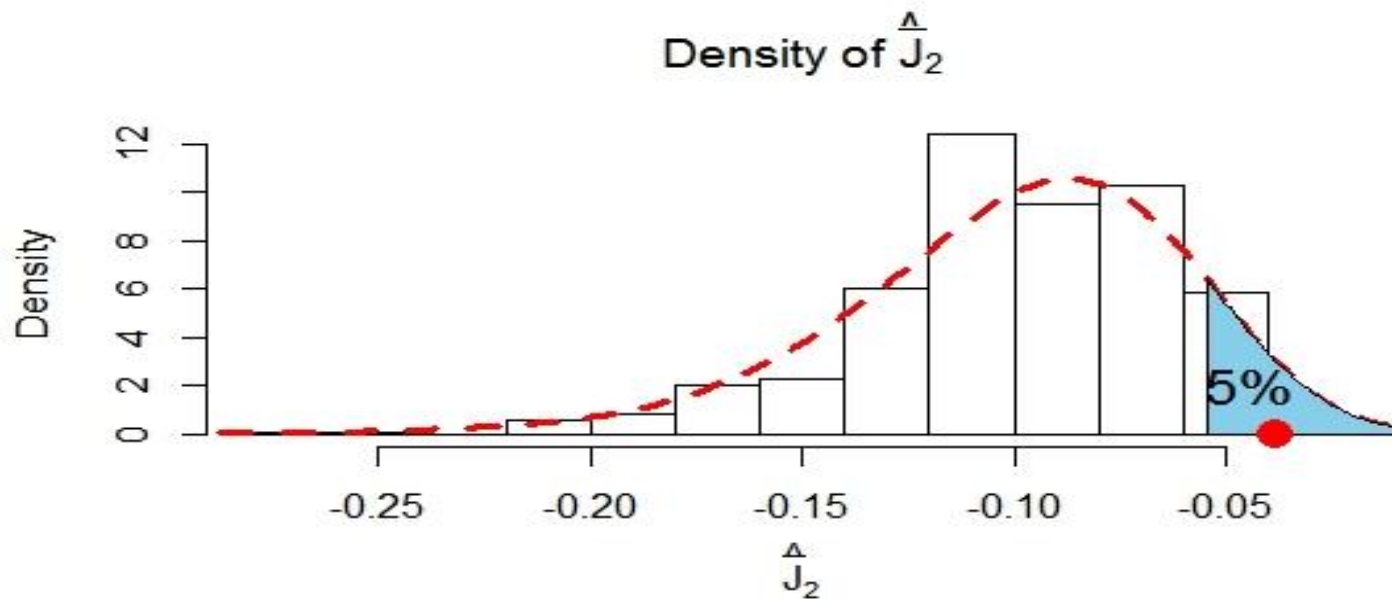
برای آزمون فرض (۲) در سطح α ، چندک $1 - \alpha$ توزیع آماره آزمون تحت فرض را استخراج کنیم و آن را با مقدار آماره آزمون برای داده‌های مشاهده شده، مقایسه کنیم.

بررسی وجود روند نرخ باروری بر حسب سن در استرالیا



- داده‌های مربوط به نرخ باروری بر حسب سن در استرالیا برای سال‌های ۱۹۲۱ تا ۲۰۰۶ برای افراد بین ۱۵ تا ۴۹ سال (هر رنگ نشان دهنده تابع یک سال)





تابع چگالی هسته‌ای برآورد شده از آماره‌های شبیه‌سازی شده
در آزمون (۲)

سطح معنی داری	۰/۱	۰/۰۵	۰/۰۱
آماره آزمون، آزمون (۲)	-۰/۰۳۸	-۰/۰۳۸	-۰/۰۳۸
مقدار چندک	-۰/۰۵۸	-۰/۰۵۴	-۰/۰۴۵

با تشکر از توجه شما