

پیش‌فرض‌های آماری، پایه بسیاری از آزمون‌های آماری تک متغیری و چندمتغیری است. شرایط مهم و اساسی برای تحلیل داده‌های چندمتغیری، برقراری پیش‌فرض‌های نرمال بودن، خطی بودن و یکسانی پراکندگی داده‌هاست. چنانچه یک یا چندتا از این مفروضه‌ها نادیده گرفته شود، در این صورت، در نتایج آماری سوگیری یا تحریف رخ می‌دهد (میزر و همکاران، ۱۳۹۱: ۱۰۷). قبل از انجام تحلیل‌های آماری تک متغیره و چندمتغیره (که پیش‌فرض‌های نرمال بودن، خطی بودن، یکسانی پراکندگی و نبود هم خطی چندگانه بین متغیرهای مستقل در مورد آن‌ها باید صدق کند) باید برقراری پیش‌فرض‌های آماری را بیازماییم.

اگر انحراف از پیش‌فرض‌های آماری ناچیز باشد می‌توان با کمی تسامح و تساهل این انحراف را نادیده گرفت و به ادامه تحلیل پرداخت. اگر انحراف از پیش‌فرض‌ها قابل توجه باشد باید یا از روش تبدیل داده‌ها برای برقرارکردن مجدد پیش‌فرض‌ها استفاده کنیم یا از آزمون‌های جایگزین استفاده کنیم که پیش‌فرض‌های فوق را مطرح نمی‌کنند (آزمون‌های ناپارامتریک).

در ادامه به توضیح چهار پیش‌فرض آماری نرمال بودن، خطی بودن رابطه‌ها، یکسانی پراکندگی و نبود هم خطی چندگانه می‌پردازیم و روش آزمون هر کدام را توضیح می‌دهیم.

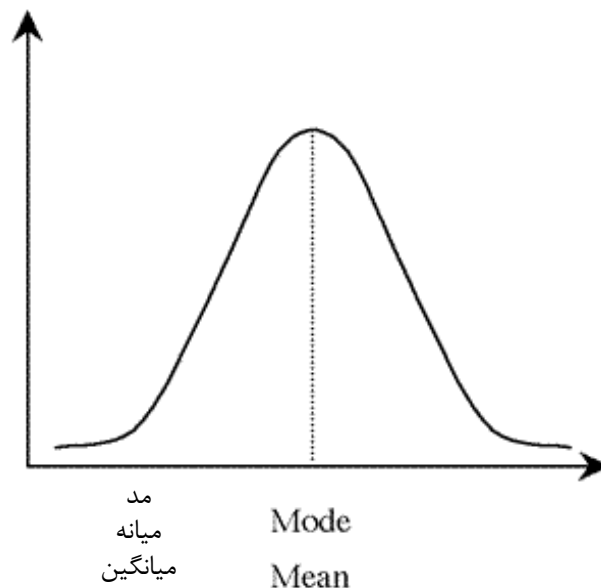
آزمون نرمال بودن داده‌ها

مفهوم توزیع نرمال در مورد داده‌های پارامتری صدق می‌کند (نه داده‌های ناپارامتری). آزمون نرمال بودن، با ایجاد یک نمودار احتمال نرمال بودن (به شکل زنگوله و نیز متقارن نسبت به میانگین)، به آزمون این فرض می‌پردازد که آیا مشاهدات پژوهش از توزیع نرمال تبعیت می‌کنند یا خیر. بسیاری از مشخصه‌های انسانی مانند هوش، نگرش‌ها و شخصیت در جمعیت (جامعه) دارای توزیع نسبتاً نرمال هستند. البته توزیع نرمال به معنای توزیع استاندارد یا مطلوب نیست.

نرمال بودن، اساسی‌ترین پیش‌فرض تحلیل چندمتغیره است. اگر این فرض برقرار نباشد، برخی آزمون‌های آماری مشخص، غیرمعتبر بوده و قابل استفاده نیستند (هیر و دیگران، ۲۰۱۰). اهمیت آشنایی و سنجش نرمال بودن توزیع داده‌ها در این است که برخی از روش‌های آماری مانند همبستگی پیرسون، آزمون‌های t و آزمون تحلیل واریانس بر فرض نرمال بودن توزیع داده‌ها (در جامعه) استوارند. همچنین برآورد پارامتر جمعیت نیز با اتکاء به نرمال بودن توزیع متغیر در جمعیت صورت می‌گیرد.

توزیع نرمال دارای ویژگی‌های زیر است:

- ۱- متقارن بوده، حداکثر ارتفاع در میانگین قرار دارد. نیمی از نمره‌ها در بالای میانگین و نیمی دیگر در پایین میانگین قرار دارد.
- ۲- مقادیر نما، میانه و میانگین برابر است.
- ۳- منحنی توزیع شبیه زنگوله است (ساعی، ۱۳۸۸: ۷۹)



شکل ۳-۴- منحنی توزیع نرمال

☑ نکته: چنانچه توزیع یک متغیر در جمعیت نرمال باشد و ما از این جمعیت نمونه‌ای با حجم ۳۰ یا بیشتر ($n \geq 30$) به طور تصادفی ساده انتخاب کنیم، در این صورت توزیع متغیر در نمونه ما نیز نرمال خواهد بود.

برای تشخیص وضعیت توزیع داده‌ها (وضعیت نرمال بودن یا بررسی چولگی و کشیدگی) روش‌های متعددی وجود دارد. در جدول ۲-۳ برخی روش‌های تشخیص وضعیت نرمال بودن توزیع داده‌ها نشان داده شده است.

جدول ۲-۳- روش‌های عددی و تصویری ارزیابی وضعیت توزیع داده‌ها (آزمون‌های سنجش نرمال بودن)

انواع روش‌ها		ماهیت روش‌ها
روش‌های عددی	روش‌های گرافیکی	
کشیدگی کجی	نمودار ساقه و برگ نمودار جعبه‌ای نمودار هیستوگرام	توصیفی
آزمون کولموگروف-اسمیرنوف آزمون شاپیرو-ویلک	نمودار احتمال - احتمال (P-P) نمودار چارک - چارک (Q-Q)	مبتنی بر نظریه

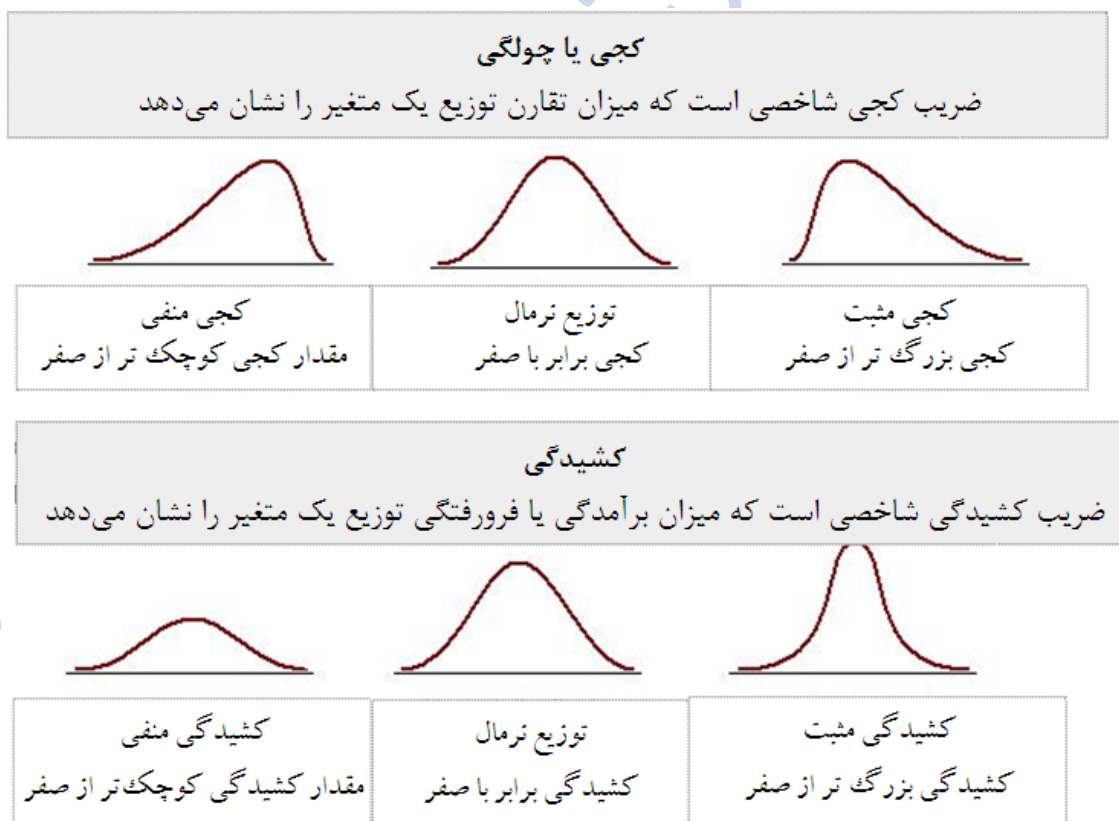
(اقتباس از حبیب پور و صفری، ۱۳۸۸: ۴۰۸)

در این کتاب از بین روش‌های ذکر شده در جدول ۳-۲، از روش‌های عددی کجی و کشیدگی و آزمون‌های کولموگروف-اسمیرنوف و شاپیرو-ویلک بهره می‌گیریم.

۱) کجی و کشیدگی

کجی یا چولگی در آمار، مقداریست از تقارن توزیع یک متغیر در اطراف میانگین. در واقع کجی، انحراف منحنی نرمال از حالت تقارن است. مقدار کجی می‌تواند مثبت یا منفی باشد. در حالت کجی مثبت، میانگین بزرگتر از میانه و میانه بزرگتر از مد است و در حالت کجی منفی، مد بزرگتر از میانه و میانه بزرگتر از میانگین است.

کشیدگی بیان‌کننده نحوه انباشته‌شدن نمره‌ها در مرکز توزیع یک متغیر است. در واقع کشیدگی، برآمدگی یا فرورفتگی منحنی توزیع نرمال است. در حالت کشیدگی تقارن توزیع حفظ می‌شود و دو نیمه منحنی متقارن هستند، اما نقطه اوج منحنی نرمال دچار تغییر می‌شود. مقدار کشیدگی می‌تواند مثبت یا منفی باشد. در حالت کشیدگی منفی، منحنی توزیع متغیر فرورفته‌تر از حالت نرمال می‌شود و نقطه اوج توزیع متغیر، پایین‌تر از توزیع نرمال است. در حالت کشیدگی مثبت، منحنی توزیع متغیر برآمده‌تر از منحنی نرمال می‌شود و نقطه اوج توزیع متغیر، بالاتر از توزیع نرمال است. در شکل شماره ۳-۵ منحنی توزیع نرمال و انواع حالت‌های کشیدگی و چولگی نمایش داده شده است.



شکل ۳-۵- مقایسه منحنی توزیع نرمال با منحنی‌های دارای کجی و کشیدگی

در مورد کجی و کشیدگی، برخی آمارشناسان تفسیر ± 1 را در مورد کجی، کشیدگی یا هر دو ترجیح می‌دهند (میزر و دیگران، ۱۳۹۱: ۸۵). به این معنا که چنانچه مقدار کجی (تقارن توزیع) و کشیدگی در دامنه -1 تا $+1$ نباشد (یعنی یا بزرگتر از $+1$ و یا کوچکتر از -1 باشد)، به عنوان شاخص مهمی برای انحراف از نرمال بودن تلقی می‌شود.

۲) آزمون‌های آماری نرمال بودن

آزمون‌های آماری جهت سنجش نرمال بودن شامل آزمون کولموگروف-اسمیرنوف^۱ و آزمون شاپیرو-ویلک^۲ است. اگر چه این هر دو آزمون می‌توانند به طور مؤثری به کار گرفته شوند، استیونس (۲۰۰۲) استفاده از آزمون شاپیرو-ویلک را پیشنهاد می‌کند، زیرا به نظر می‌رسد «در تشخیص انحراف از نرمال قوی‌تر است». معنی‌داری آماری این شاخص‌ها به طور آرمانی در سطح آلفای $(P < .001)$ ، بیانگر تخطی از نرمال بودن تک متغیری است (میزر، گامست و گارینو، ۱۳۹۱: ۱۰۷). یعنی اگر معنی‌داری این آزمون‌ها کمتر از مقدار $.001$ به دست بیاید نشان می‌دهد که توزیع متغیر نرمال نیست و اگر سطح معنی‌داری بیشتر از $.001$ شود، نشان می‌دهد توزیع متغیر نرمال یا نزدیک به نرمال است.

☑ نکته: آزمون‌های نرمال بودن مختص متغیرهای کمی (فاصله‌ای/نسبی) است.

مثال

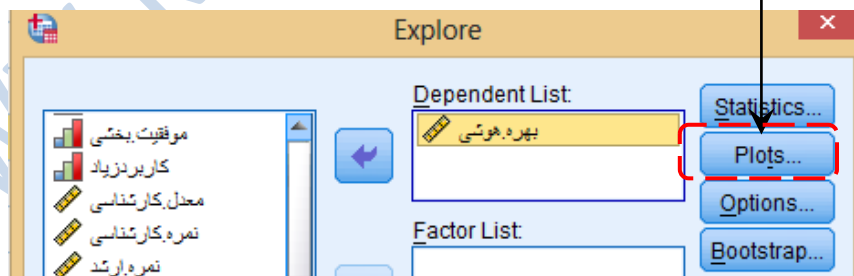
می‌خواهیم نرمال بودن توزیع متغیر بهره هوشی دانشجویان را بررسی کنیم.

اجرا:

مسیر زیر را دنبال می‌کنیم:

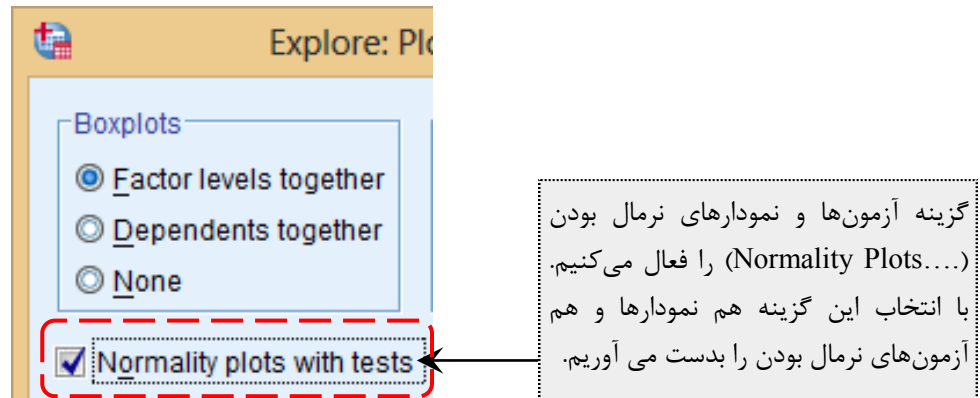
Analyze ---> Descriptive Statistics ---> Explore

در کادر Dependent List متغیر بهره هوشی که می‌خواهیم نرمال بودن آن را بیازماییم وارد می‌کنیم. بر روی گزینه Plots (نمودارها) کلیک می‌کنیم.



¹ Kolmogorov-Smirnov

² Shapiro-Wilk



نتایج:

در جدول بعد نتیجه آزمون نرمال بودن داده‌ها گزارش شده است. همان‌طور که مشاهده می‌شود برنامه دو آزمون نرمال بودن را ارائه می‌دهد: آزمون شاپیرو-ویلک (Shapiro-Wilk) و آزمون کولموگروف-اسمیرنوف (Kolmogorov-Smirnov).

جهت بررسی نرمال بودن داده‌ها، میزان Sig (سطح معنی‌داری) متناظر با آن‌ها را بررسی می‌کنیم. چنانچه میزان Sig برای هرکدام از متغیرها بیشتر از ۰.۰۱ باشد، نتیجه می‌گیریم که داده‌ها از توزیع نرمال برخوردارند (به بیان صحیح‌تر، توزیع داده‌ها از توزیع نرمال انحراف قابل توجهی ندارد).

☑ نکته: بیشتر نویسندگان محتاط‌ترند و معتقدند برای این که نرمال بودن داده‌ها را بپذیریم؛ سطح معنی‌داری آزمون‌های کولموگروف-اسمیرنوف و شاپیرو-ویلک باید بیشتر از ۰.۰۵ باشد ($P > 0.05$).

☑ نکته: آزمون‌های کولموگروف-اسمیرنوف و شاپیرو-ویلک به انحراف از توزیع نرمال بسیار حساس هستند؛ در نتیجه در نمونه‌های با حجم زیاد، انحراف کوچکی از توزیع نرمال ممکن است موجب شود این آزمون‌ها توزیع متغیرها را غیر نرمال نشان دهند. در نتیجه زمانی که حجم نمونه زیاد است (حدوداً بیشتر از ۱۰۰)، بهتر است از روش‌های دیگر بررسی نرمال بودن هم استفاده شود. از جمله این روش‌ها می‌توان به شاخص‌های کجی و کشیدگی و نمودارهایی مانند هیستوگرام، ساقه و برگ و نمودار نرمال بودن Q-Q اشاره کرد.

نتایج آزمون شاپیرو-ویلک نشان می‌دهد که توزیع متغیر بهره هوشی در نمونه آماری ما نرمال نیست ($P < 0.01$). سطح معنی‌داری به دست آمده کمتر از مقدار ۰.۰۱ است و نشان می‌دهد که توزیع متغیر بهره هوشی در نمونه نرمال نیست.

Tests of Normality

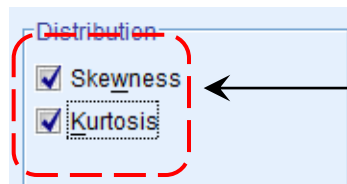
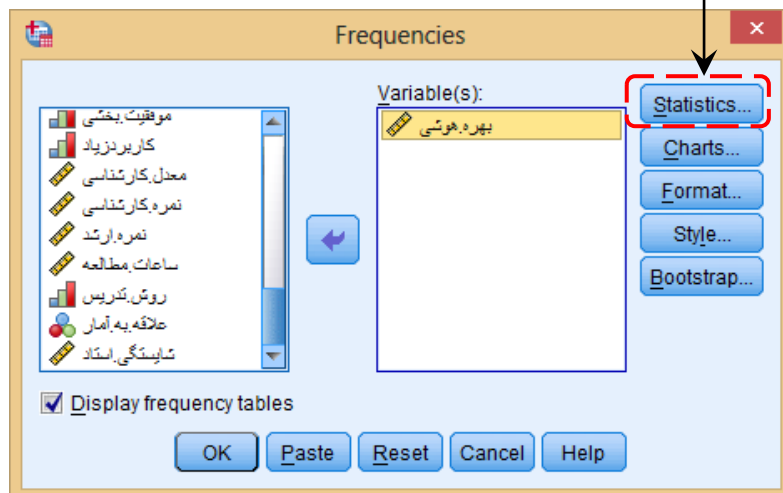
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
	آماره	درجه آزادی	سطح معنی‌داری	آماره	درجه آزادی	سطح معنی‌داری
IQ	.181	100	.000	.636	100	.000

اجرا:

قصد داریم نرمال بودن متغیر بهره هوشی را با شاخص‌های کجی و کشیدگی آزمون کنیم. مسیر زیر را دنبال می‌کنیم:

Analyze ---> Descriptive Statistics ---> Frequencies

در کادر Variables متغیری که می‌خواهیم کجی و چولگی آن را محاسبه کنیم (بهره هوشی) وارد می‌کنیم. سپس گزینه Statistics را انتخاب می‌کنیم.



در کادر شکل توزیع (Distribution) گزینه‌های کجی (Skewness) و کشیدگی (Kurtosis) را انتخاب می‌کنیم. گزینه Continue و OK را انتخاب می‌کنیم.

نتایج:

جدول زیر مقادیر شاخص‌های کجی و کشیدگی را نشان می‌دهد. همان طور که توضیح داده شد، مقادیر کجی و کشیدگی باید بین -1 تا $+1$ باشد تا بتوانیم نرمال بودن توزیع متغیر را بپذیریم. مقدار کجی به دست آمده برابر با -1.32 است که کمتر از مقدار -1 (منفی یک) است و نشان از این دارد که توزیع متغیر با توزیع نرمال فاصله دارد. مقدار کشیدگی هم برابر با 26.69 است که بسیار بیشتر از مقدار $+1$ (مثبت یک) است که نشان از توزیع غیرنرمال متغیر بهره هوشی دارد. توزیع متغیر دارای کجی منفی و کشیدگی مثبت است و از توزیع نرمال برخوردار نیست. پیشنهاد می‌شود یا از روش‌های ناپارامتریک برای آزمون‌های آماری استفاده شود و یا با استفاده از روش تبدیل داده‌ها، سعی شود که توزیع متغیر، نرمال یا نزدیک به نرمال شود.

Statistics

بهره هوشی

N	Valid	100
	Missing	0
Skewness (کجی)		-1.32
Std. Error of Skewness (خطای استاندارد کجی)		.24
Kurtosis (کشیدگی)		26.69
Std. Error of Kurtosis (خطای استاندارد کشیدگی)		.47

