

## آموزش رگرسیون خطی با SPSS

### تاریخچه رگرسیون

واژه «رگرسیون» نخستین بار توسط سر فرانسیس گالتون (Sir Francis Galton، ۱۸۲۲-۱۹۱۱)، دانشمند انگلیسی و پسرعموی چارلز داروین، در سال ۱۸۷۷ وارد ادبیات علمی شد. گالتون که به مطالعه وراثت صفات جسمانی علاقه‌مند بود، رابطه میان قد پدران و پسران را بررسی کرد. یافته شگفت‌انگیز او این بود که پدران بسیار بلندقد، پسرانی دارند که از میانگین جامعه بلندترند اما به اندازه خودشان بلند نیستند؛ و پدران بسیار کوتاه‌قد نیز پسرانی دارند که از میانگین کوتاه‌ترند اما نه به اندازه خود پدران. او این پدیده را بازگشت به میانگین (Regression toward the mean) نامید. به همین دلیل، روش آماری که برای مدل‌سازی چنین روابطی به کار می‌رود، «رگرسیون» خوانده شد. بعدها کارل پیرسون (Karl Pearson) و اودنی یول (Udny Yule) مبانی ریاضی رگرسیون را گسترش دادند و آن را از حیطة زیست‌سنجی به علوم رفتاری وارد کردند. امروزه رگرسیون یکی از قدرتمندترین ابزارهای پیش‌بینی و تبیین در علوم رفتاری است.

### شباهت و تفاوت رگرسیون با همبستگی

بسیاری از دانشجویان تازه‌کار، همبستگی (Correlation) و رگرسیون (Regression) را یکسان می‌پندارند، اما این دو مفهوم اگرچه مرتبط هستند، کاربردها و منطق متفاوتی دارند. فرض کنید رابطه بین ساعت خواب شب قبل از امتحان و نمره امتحان را بررسی می‌کنیم:

همبستگی به ما می‌گوید:

«آیا بین خواب و نمره رابطه وجود دارد؟ جهت آن مثبت است یا منفی؟ شدت آن چقدر است؟»

مثلاً  $r = 0.65$  (رابطه از نظر جهت مثبت و از نظر شدت متوسط تا قوی).

رگرسیون به ما می‌گوید:

«اگر یک دانشجو ۲ ساعت بیشتر از میانگین بخوابد، نمره او چند واحد افزایش می‌یابد؟» مثلاً  $b = 1.5$

یعنی هر یک ساعت خواب اضافه، نمره را ۱.۵ نمره بالا می‌برد.

یا مثال دیگر اینکه براساس معدل و رتبه کنکور تعدادی افراد ما به یک معادله یا الگویی می‌رسیم که می‌توانیم بر اساس آن رتبه کنکور احتمالی یک فرد را بر اساس معدل وی پیش‌بینی کنیم.

پس در نتیجه اگر فقط می‌خواهید بدانید دو متغیر با هم حرکت می‌کنند یا نه، از همبستگی استفاده کنید. اگر می‌خواهید پیش‌بینی کنید یا به نوعی بگویید «چقدر تأثیر» دارد، از رگرسیون استفاده کنید.

رگرسیون خطی انواع متعددی دارد. در رگرسیون خطی ساده ما یک متغیر پیش‌بین و یک متغیر وابسته داریم. در رگرسیون چندگانه ما یک متغیر وابسته و چند متغیر مستقل داریم. مثالی که در این فایل ارائه می‌دهیم چندگانه است.

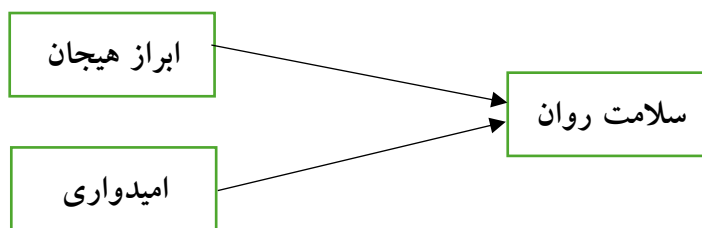
بریم سراغ یک مثال ملموس‌تر و کار با SPSS. فرض کنید عنوان پژوهش این باشد:

### پیش‌بینی سلامت روان بر اساس ابراز هیجان و امیدواری در دانشجویان دانشگاه هرمزگان

سلامت روان متغیر وابسته یا ملاک است و متغیرهای ابراز هیجان و امیدواری متغیرهای مستقل یا پیش‌بین هستند. متغیرهای مستقل رو X در نظر بگیرید و متغیر وابسته رو Y که ما توی این مثال دو تا X داریم که ما می‌خواهیم از طریق این یکس‌ها مقدار Y رو پیش‌بینی کنیم. مقیاس اندازه‌گیری متغیر وابسته حتما باید فاصله‌ای یا نسبی باشد. ما می‌خواهیم بررسی کنیم که (۱) ابراز هیجان و امیدواری بایکدیگر چقدر میتوانند سلامت روان را پیش‌بینی کنند؟ (۲) آیا مقدار پیش‌بینی و مدلی که ارائه شده معنادار است یا خیر؟ و (۳)

سهم هر متغیر پیش‌بین به صورت جداگانه چقدر است؟

شکل مدل به صورت زیر است:



فرضیه‌های این مدل به صورت زیر است:

#### فرضیه پژوهش

۱. ابراز هیجان می‌تواند سلامت روان را به صورت معنادار پیش‌بینی نماید.

۲. امیدواری می‌تواند سلامت روان را به صورت معنادار پیش‌بینی نماید.

#### فرضیه صفر

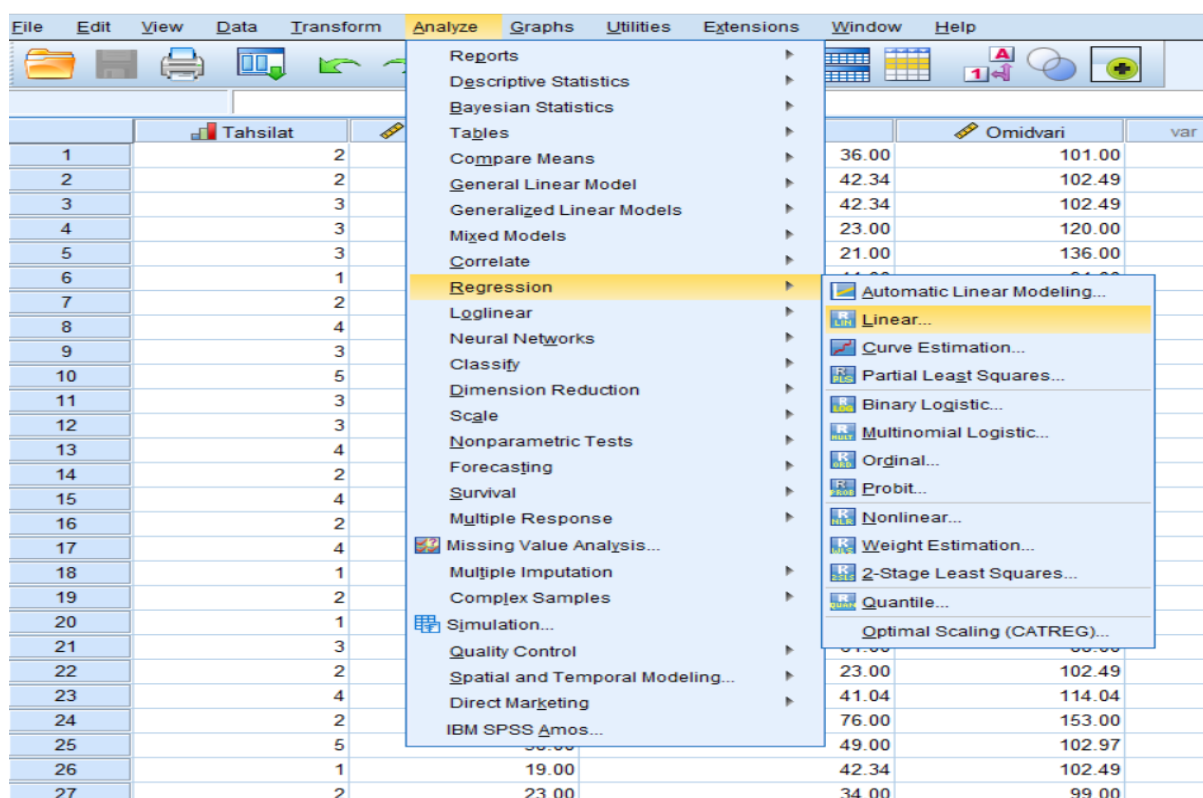
۱. ابراز هیجان نمی تواند سلامت روان را به صورت معنادار پیش بینی نماید.

۲. امیدواری نمی تواند سلامت روان را به صورت معنادار پیش بینی نماید.

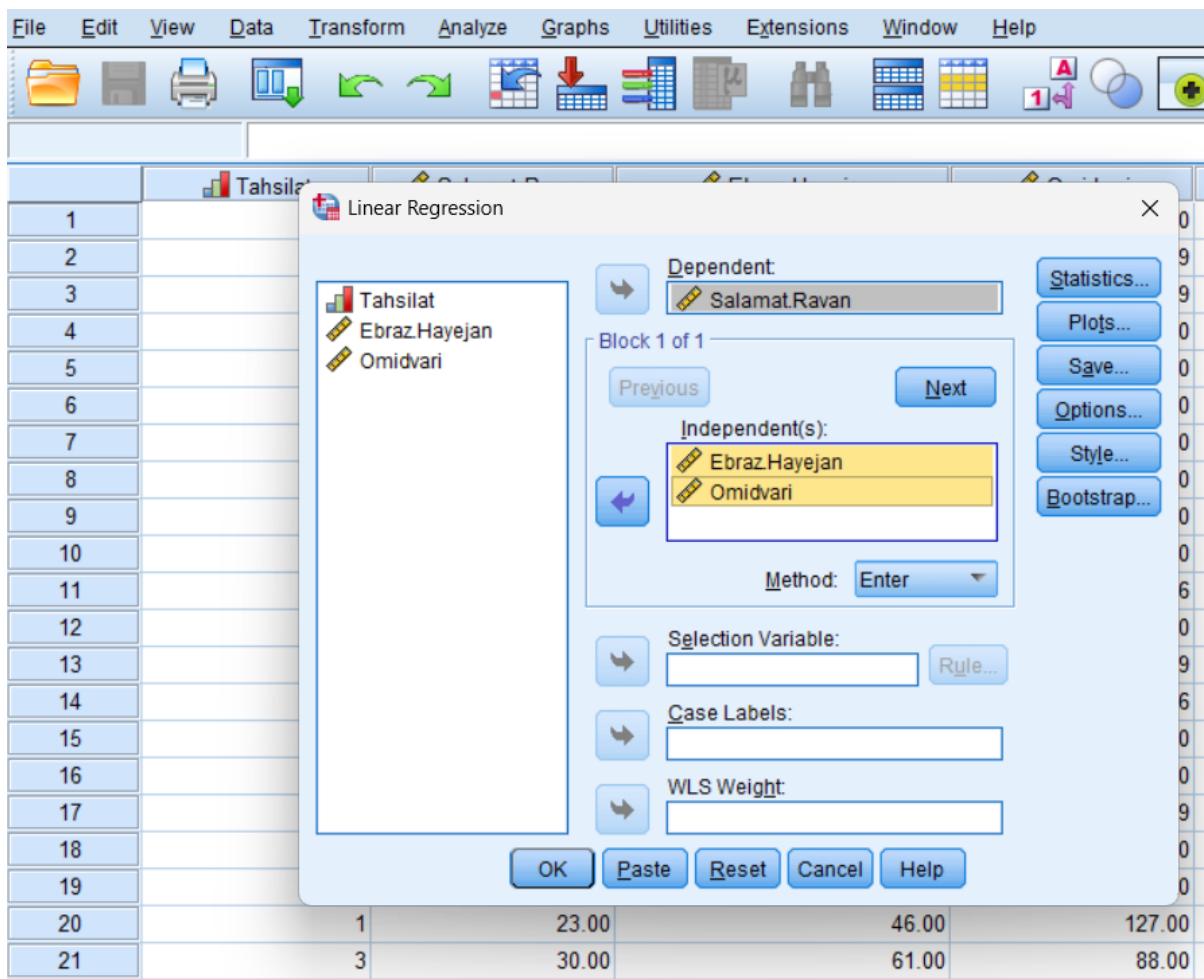
بریم سراغ SPSS

ابتدا مسیر زیر را دنبال کنید:

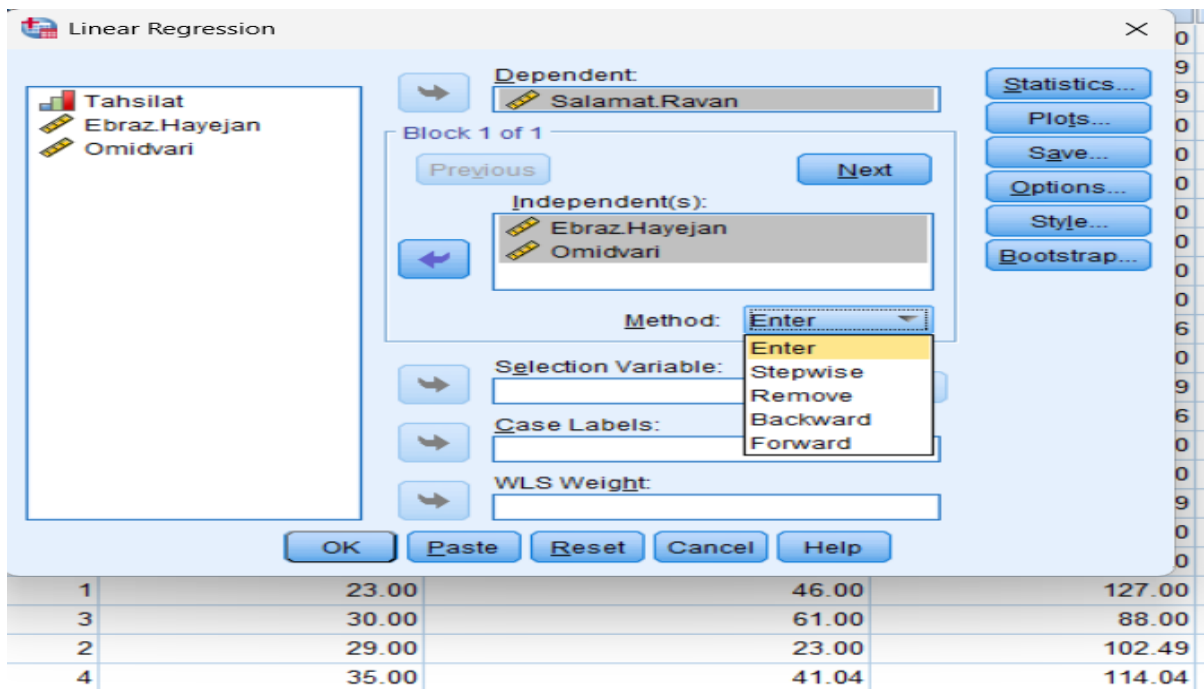
Analyze → Regression → Linear



وقتی اوکی کنید باکس زیر می آید. باکس **Dependent** برای متغیر وابسته ما است (اگر دقت کنید فقط یک متغیر میتواند به این باکس منتقل شود). در نتیجه متغیر سلامت روان را از باکس سمت چپ به آن منتقل میکنیم (از طریق فلش آبی). باکس **Independent** برای متغیرهای مستقل است. در نتیجه متغیرهای ابراز هیجان و امیدواری را از باکس سمت چپ به آن منتقل کنید.



**نکته:** برای انجام رگرسیون خطی روش‌های مختلف است. در قسمت روش (Method) انواع آن مشخص است. اگر روی باکس متد کلیک کنید، کشوی آن باز میشود و میتوانید انواع آن را ببینید. انواع آن شامل همزمان (Enter)، گام به گام (Stepwise)، حذفی (Remove)، پسرونده (Backward) و پیشرونده (Forward) است. ما بیشتر از روش همزمان استفاده میکنیم و گاهی از روش گام به گام. در روش همزمان همه متغیرها یک دفعه وارد مدل می‌شود. در روش گام به گام، نرم افزار خودش تصمیم می‌گیرد کدام متغیرها را وارد کند. اول قوی‌ترین متغیر را می‌آورد، بعد می‌بیند اگر متغیر بعدی را اضافه کند بهتر می‌شود یا نه. ما به صورت پیشفرض روش همزمان را انتخاب میکنیم. در شکل زیر میتوانید انواع آن را مشاهده نمایید:



**نکته:** برای انجام رگرسیون خطی یکسری پیشفرض هست که فعلا از بیان آن صرف نظر میکنم: خطی بودن، نرمال بودن متغیر وابسته و مقادیر خطا، عدم هم خطی، استقلال خطا. برای اطلاعات بیشتر می توانید به این مقاله مراجعه کنید: بهنام فر، رضا و راستی، اعظم (۱۳۹۴). رگرسیون خطی، نرمال بودن توزیع مقادیر خطا یا نرمال بودن توزیع متغیر وابسته؟. مجله ایرانی آموزش در علوم پزشکی، ۱۵: ۲۶۵-۲۶۳.

<http://ijme.mui.ac.ir/article-1-3780-fa.html>

اگر اوکی را کلیک کنید خروجی زیر حاصل می شود:

Model	Variables Entered	Variables Removed	Method
1	Omidvari, Ebraz.Hayejan <sup>b</sup>		Enter

a. Dependent Variable: Salamat.Ravan  
b. All requested variables entered.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.394 <sup>a</sup>	.155	.146	8.82352

a. Predictors: (Constant), Omidvari, Ebraz.Hayejan

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2811.097	2	1405.548	18.054	.000 <sup>b</sup>
	Residual	15337.344	197	77.855		
	Total	18148.441	199			

a. Dependent Variable: Salamat.Ravan  
b. Predictors: (Constant), Omidvari, Ebraz.Hayejan

Model		Unstandardized Coefficients		Standardized Coefficients		t	Sig.
		B	Std. Error	Beta			
1	(Constant)	13.028	3.033			4.295	.000
	Ebraz.Hayejan	.130	.046	.188		2.846	.005
	Omidvari	.120	.025	.321		4.860	.000

a. Dependent Variable: Salamat.Ravan

در خروجی‌ها، جداولی که مهم است و نیازحتمی به گزارش و تفسیر اون هست شامل این جداول است: خلاصه مدل (Model Summary)، آنوا (ANOVA)، و ضرایب (Coefficients).

حالا تک تک اونها رو تحلیل میکنیم:

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.394 <sup>a</sup>	.155	.146	8.82352

جدول خلاصه مدل میزان همبستگی چندگانه (R)، میزان پیش‌بینی (R Square)، پیش‌بینی تعدیل شده (Adjusted R Square) و خطای استاندارد برآورد (Std. Error of the Estimate) را نشون میده. اونهایی که مهم است رو هایلایت کردم و سعی میکنم همین‌ها رو خلاصه توضیح بدم. R چی میگه: می‌گه مقدار همبستگی همزمان متغیرهای پیش‌بین ابراز هیجان و امیدآوری با هم با متغیر وابسته سلامت روان ۰.۳۹۴ است. مقدار پیش‌بینی سلامت روان براساس متغیرهای پیش‌بین ابراز هیجان و امیدآوری با هم ۰.۱۵۵ است که اگر در ۱۰۰ ضرب شود برابر است با ۱۵.۵. یعنی در این مدل حدود ۱۵.۵ درصد از سلامت روان توسط ابراز هیجان و امیدآوری پیش‌بینی می‌شود. (این همان جواب سوال ۱ است که در بالا زیر آن خط کشیده شده است)

**نکته:** در مورد تفاوت این  $r$  با  $r$  همبستگی پیرسون، باید بگم اگر همبستگی پیرسون را از مسیر correlation در SPSS بگیرید، همبستگی پیرسون میزان رابطه یا همبستگی متغیرها را دو به دو نشون میده (مثلا میزان همبستگی امیدواری با سلامت روان، یا میزان همبستگی ابراز هیجان با سلامت روان) اما اینجا به صورت همزمان میزان همبستگی ابراز هیجان و امیدآوری را با سلامت روان نشان می‌دهد. مورد دوم اینکه اگر سلامت روان را به ۱۰۰ قسمت تقسیم کنید، این ۲ متغیر فقط ۱۵.۵ درصد اون رو پیش‌بینی می‌کنند؛ از جهت دیگر ۸۴.۵ درصد از عواملی که روی سلامت روان تأثیر می‌گذارند، چیزهای دیگری هستند که در این مدل بررسی نشده‌اند (مثل استرس، حمایت خانواده، وضعیت اقتصادی، ژنتیک و ...).

بریم سراغ جدول بعدی. جدول زیر مقدار آنوا را نشان می‌دهد و این جدول برای پاسخگویی به سوال دوم است که در بالا زیر آن خط کشیده شده است:

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2811.097	2	1405.548	18.054	.000 <sup>b</sup>
	Residual	15337.344	197	77.855		
	Total	18148.441	199			

جدول آنوا به این پاسخ می‌دهد که آیا مقدار پیش‌بینی و مدلی که ارائه شده معنادار و قابل استناد است یا خیر. میزان F برابر با ۱۸.۰۵۴ است و مقدار سطح معناداری آن (Sig.) برابر با ۰.۰۰۰ است. از آنجایی که سطح معناداری کمتر از ۰.۰۵ (پنج صدم) است در نتیجه مقدار به دست آمده معنادار است و نتیجه گرفته می‌شود که مقدار پیش‌بینی و مدل ارائه شده معنادار است.

نکته: اگر مقدار سطح معناداری بزرگتر از ۰.۰۵ باشد؛ مثلاً ۰.۰۵۹ بود یعنی معنادار نبود. پس در تفسیر سطح معناداری اشتباه نکنید. چون برخی از دانشجویان در تفسیر ۰.۰۵ اشتباه می‌کنند، بدانید که این ۰.۰۵ همان ۰.۰۵۰ (پنجاه هزارم) است. پس ببینید مقدار به دست آمده از پنجاه هزارم بزرگتر است یا خیر.

جدول بعدی جدول ضرایب است و برای پاسخگویی به سوال سوم است.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.028	3.033		4.295	.000
	Ebraz.Hayejan	.130	.046	.188	2.846	.005
	Omidvari	.120	.025	.321	4.860	.000

این جدول به شما می‌گوید هر کدام از متغیرهای مستقل به تنهایی چقدر پیش‌بینی می‌کند یا به نوعی چقدر تأثیر دارد و نقش هر کدام به تنهایی معنادار است یا خیر. پس در نتیجه ممکن است برخی متغیرها نقش معناداری به تنهایی نداشته باشند.

Beta ضرایب استاندارد رو نشون میده. مقدار بتا یا ضریب استاندارد ابراز هیجان برابر با ۰.۱۸۸ است و از آنجایی که سطح معناداری آن ۰.۰۰۵ است و از ۰.۰۵ کمتر است در نتیجه این مقدار معنادار است. در نتیجه فرضیه پژوهشی اول تایید و فرضیه صفر اول رد می‌شود و نتیجه گرفته می‌شود که ابراز هیجان می‌تواند به

صورت معنادار سلامت روان را پیش‌بینی نماید. در نتیجه انتظار می‌رود با افزایش یک نمره در ابراز هیجان به میزان ۰.۱۸۸ واحد سلامت روان افزایش یابد.

مقدار ضریب استاندارد امیدواری برابر با ۰.۳۲۱ است و از آنجایی که سطح معناداری آن ۰.۰۰۰ است و از ۰.۰۵ کمتر است در نتیجه این مقدار معنادار است. در نتیجه فرضیه پژوهشی دوم تایید و فرضیه صفر دوم رد می‌شود و نتیجه گرفته می‌شود که امیدواری می‌تواند به صورت معنادار سلامت روان را پیش‌بینی نماید. در نتیجه انتظار می‌رود با افزایش یک نمره در امیدواری به میزان ۰.۳۲۱ واحد سلامت روان افزایش یابد.

**نکته:** مقدار بتا یا ضریب استاندارد در اینجا مثبت است اما ممکن است منفی هم باشد. اگر منفی شد عکس این نتیجه صادق است و نشان می‌دهد وقتی متغیر مستقل افزایش یابد متغیر وابسته کاهش پیدا خواهد کرد و بالعکس.

**نکته:** دانشجویان عزیز دقت فرمایید، قسمت‌هایی که با هایلایت زرد مشخص شده از مواردی هستند که در امتحان از شما درخواست می‌شود تا آن‌ها را بنویسید.

**دکتر هادی سلیمی**

عضو هیات علمی گروه مشاوره و روان‌شناسی دانشگاه هرمزگان

hadisalimi69@gmail.com hadisalimi69@yahoo.com h.salimi@hormozgan.ac.ir

\* استفاده از این فایل برای آموزش، آزاد است و هیچ منعی ندارد.