

با سلام و احترام،

## ادامه فصل بیستم و چهارم: رگرسیون خطی (Linear Regression)

خروجی‌های رگرسیون

### توصیف‌های رگرسیون

جدول آماره‌های توصیفی هر متغیر را نشان می‌دهد و شامل میانگین، انحراف استاندارد و تعداد موارد مقادیر معتبر می‌باشد.

	Mean	Std. Deviation	N
Average female life expectancy	70.01	10.62	107
Fertility: average number of kids	3.563	1.902	107
Infant mortality (deaths per 1000 live births)	42.937	38.153	107

### همبستگی‌های رگرسیون

در جدول همبستگی‌ها ضرائب همبستگی پیرسون، سطح معنی‌داری، و تعداد موارد با مقادیر معتبر وجود دارد. نکات بیان شده درباره ضریب پیرسون، سطح معنی‌داری و **N** در اینجا نیز صادق است.

### متغیرهای وارد شده / حذف شده رگرسیون

این جدول متغیرهایی را که در هر مرحله وارد و یا حذف شده‌اند نشان می‌دهد. متغیرهایی که در مدل در هر مرحله‌ای وارد شده‌اند در ستون **Variables Entered** فهرست شده‌اند. متغیرهای حذف شده از مدل در هر مرحله در ستون **Variable Removed** فهرست شده‌اند. در ستون **Method** روش انتخاب شده برای حذف یا وارد کردن متغیر نمایش داده شده است.

### خلاصه مدل رگرسیون

جدول شامل **R**، ضریب تعیینی، ضریب تعیینی تعدیل شده و خطای استاندارد می‌باشد (ستون دوم تا پنجم). **R** یعنی ضریب همبستگی چندگانه، همبستگی بین مقادیر مشاهده شده و پیش‌بینی شده متغیر وابسته است. مقادیر **R** مربوط به مدل تولید شده با دستورالعمل رگرسیون در محدوده 0 تا 1 هستند. مقادیر بزرگتر **R** روابط قوی‌تر را نشان می‌دهند.

**R Squared** نسبت تغییر در متغیر وابسته می‌باشد که با مدل رگرسیون بیان شده است. مقادیر ضریب تعیینی (مربع **R**) در محدوده 0 تا 1 هستند. مقادیر کوچک نشان می‌دهد که مدل به خوبی با داده‌ها منطبق نیست. نمونه مربع **R** گرایش دارد که نحوه انطباق مدل‌ها با جمعیت آماری را برآورد نماید.

**Adjusted R Squared** سعی دارد که مربع **R** را تصحیح نماید تا بیشترین میزان انطباق مدل را در جمعیت انعکاس دهد. از ضریب تعیینی برای تعیین اینکه کدام مدل بهتر است استفاده نمایید. یک مدل با مقدار بزرگ ضریب تعیینی را برگزینید که شامل تعداد زیادی متغیر نیست. مدل‌های با متغیرهای زیاد اغلب بیش از حد انطباق دارند و برای تفسیر کردن سخت می‌باشند.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.962	.925	.924	2.93
2	.965	.931	.930	2.82
3	.967	.935	.933	2.75
4	.971	.942	.940	2.62
5	.973	.947	.944	2.51
6	.973	.946	.944	2.52

## رگرسیون ANOVA

این جدول نتایج یک تحلیل واریانس را خلاصه کرده است. مجموع مربعات، درجات آزادی و مربع میانگین برای دو مورد از منابع تغییر، یعنی **Regression** و **Residual** از محتویات این جدول است. خروجی **Regression** اطلاعات راجع به تغییر که برای مدل شما در نظر گرفته شده است را نشان می‌دهد. خروجی **Residual** اطلاعات راجع به تغییر که برای مدل شما در نظر گرفته نشده است را نشان می‌دهد و خروجی **Total** جمع اطلاعات مربوط به **Regression** و **Residual** می‌باشد.

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11101.959	2	5550.980	684.793	.000
	Residual	843.031	104	8.106		
	Total	11944.991	106			

یک مدل با مجموع مربعات رگرسیون بزرگ در مقایسه با مجموع مربعات باقی مانده مشخص می کند که مدل برای بسیاری از تغییرات در متغیر وابسته در نظر گرفته شده است. مقادیر خیلی بزرگ مجموع مربعات باقی مانده نشان می دهند که مدل در توضیح بسیاری از تغییرات در متغیر وابسته ناتوان است و شما ممکن است بخواهید که به دنبال عوامل اضافی بگردید که در به حساب آوردن یک نسبت بزرگتر تغییر در متغیر وابسته به شما کمک می کند.

مربع میانگین مجموع مربعات تقسیم بر درجات آزادی می باشد. آماره **F** مربع میانگین رگرسیون (**MSR**) تقسیم بر مجموع مربعات باقی مانده (**MSE**) می باشد. درجات آزادی رگرسیون، صورت **df** و درجات آزادی باقی مانده، مخرج **df** برای آماره **F** می باشد. تعداد کل درجات آزادی برابر تعداد موارد منهای **1** می باشد.

اگر سطح معنی داری آماره **F** کوچک باشد (کمتر از **0.05**) آنگاه متغیرهای مستقل به خوبی تغییر در متغیر وابسته را تبیین می کنند. در صورتی که این مقدار بزرگتر از **0.05** باشد، آنگاه متغیرهای مستقل، تغییر در متغیر وابسته را تبیین نمی کنند.

## ضرایب رگرسیون

ضرایب استاندارد نشده، ضرایب مدل رگرسیون برآورد شده می باشند. در این مثال، مدل تخمینی، میزان متوسط عمر خانمها برابر **0.240 Infant mortality - 82.677 Fertility** می باشد. اغلب متغیرهای مستقل در واحدهای مختلف سنجش می شوند. ضرایب استاندارد شده یا بتا (**Beta**) برای سنجش پذیرتر کردن ضرایب رگرسیون می باشد.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	82.677	.626		132.176	.000
	Fertility	-.662	.263	-.119	-2.518	.013
	Infant mortality	-.240	.013	-.863	-19.326	.000

## شیوه‌های تشخیصی بر اساس مورد رگرسیون

این جدول آماره‌های موارد خاص را که با تعداد موردشان تعیین می‌شوند نشان می‌دهد. آماره‌ها بیشتر برای موارد با باقیمانده‌های استاندارد نشان داده شده است تا برای تعداد معینی از انحراف معیارهای دور از میانگین. باقی‌مانده‌های استاندارد شده، باقیمانده‌های معمولی تقسیم شده بر انحراف معیار نمونه باقی‌مانده‌ها می‌باشند. باقی‌مانده‌های استاندارد شده دارای میانگین 0 و انحراف معیار 1 هستند.

مقادیر مشاهده شده متغیر وابسته نمایش داده شده‌اند (Female Life Expectancy). ممکن است که مقدار پیش‌بینی شده و باقی‌مانده هر مورد را هم ببینید. باقی‌مانده (Residual) تفاوت بین مقدار مشاهده شده متغیر وابسته و مقدار پیش‌بینی شده است.

Case Number	Country	Std. Residual	Female Life Expectancy	Predicted Value	Residual
53	Iraq	2.049	68	62.84	5.16
60	Kuwait	-2.250	78	83.66	-5.66
63	Liberia	-2.359	57	51.06	5.94
97	Tanzania	-2.430	45	51.12	-6.12
103	Uganda	-2.716	43	49.81	-6.81
109	Zambia	-4.199	45	55.57	-10.57

## آماره‌های باقی‌مانده رگرسیون

این جدول آماره‌های باقی‌مانده‌ها و مقادیر پیش‌بینی شده را نشان می‌دهند. برای هر مورد، مقدار پیش‌بینی شده عبارتست از مقدار پیش‌بینی توسط مدل رگرسیون (ردیف Predicted Value) و برای هر مورد، Residual تفاوت بین مقدار مشاهده شده متغیر وابسته و مقدار پیش‌بینی شده توسط مدل است. باقی‌مانده‌ها برآوردهای خطاهای درست در مدل هستند. اگر مدل برای داده‌ها مناسب باشد، باقی‌مانده‌ها باید از یک توزیع نرمال تبعیت نمایند.

مقادیر پیش‌بینی استاندارد شده مقادیری هستند که میانگین 0 و انحراف معیار 1 دارند. (ردیف Std. Predicted Value) به طور مشابهی، باقی‌مانده‌های استاندارد، باقی‌مانده‌های معمولی تقسیم بر انحراف معیار نمونه باقی‌مانده‌ها هستند و میانگین 0 و انحراف معیار 1 دارند. کمترین، بیشترین، میانگین، انحراف معیار و اندازه نمونه موارد فوق در جدول نمایش داده شده‌اند.

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	40.00	83.66	70.02	10.34	107
Residual	-10.57	5.94	-1.09E-02	2.46	107
Std. Predicted Value	-2.892	1.325	.007	.998	107
Std. Residual	-4.199	2.359	-.004	.977	107

## همبستگی های ضریب رگرسیون

این جدول، ماتریس های همبستگی و کوواریانس متغیرهای مستقل را که در هر مرحله در مدل هستند نشان می دهد.

Model		Infant mortality	%Urban	Fertility
1	Correlations	Infant mortality		
	Covariances	Infant mortality		
2	Correlations	Infant mortality	.72796	
		%Urban	.72796	1.00000
	Covariances	Infant mortality	.00011	.00013
		%Urban	.00013	.00027
3	Correlations	Infant mortality	.46426	.78814
		%Urban	.46426	.78814
		Fertility	.00167	.00080
	Covariances	Infant mortality	.00021	.00011
		%Urban	.00011	.00026
		Fertility	.00011	.00014
				.06456

در ماتریس های همبستگی، مقادیر ضرائب همبستگی در محدوده **-1** تا **1** هستند. علامت ضریب همبستگی، جهت رابطه را (مثبت یا منفی) مشخص می نماید.

مقدار مطلق ضریب همبستگی، شدت رابطه را تعیین می نماید، مقادیر مطلق بزرگ تر رابطه های قوی تر را نشان می دهند. ضرائب همبستگی در قطر اصلی همیشه **1** هستند، چون هر متغیر یک رابطه خطی کاملاً مثبت با خودش دارد. همبستگی های بالای قطر اصلی، در پایین قطر اصلی تکرار شده اند. در ماتریس های کوواریانس، واریانس ها در بالا و پایین قطر اصلی نمایش داده شده اند. کوواریانس های بالای قطر اصلی معکوس کوواریانس های زیر قطر اصلی هستند.

Model		Infant mortality	%Urban	Fertility
1	Covariances	Infant mortality	.00006	
2	Covariances	Infant mortality	.00011	.00013
		%Urban	.00013	.00027
3	Covariances	Infant mortality	.00021	.00011
		%Urban	.00011	.00026
		Fertility	.00011	.00014
				.06456

## تشخیص هم خطی بودن رگرسیون

این جدول آماره‌هایی را نشان می‌دهد که با آن می‌توانید مشکلات مربوط به هم‌خطی بودن را تعیین نمایید. هم‌خطی (یا هم‌خطی چندگانه) وضعیت نامطلوبی است که در جایی که همبستگی‌های متغیرهای رشته‌ای هستند ایجاد می‌شود.



Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Infant Mortality	%Urban
1	1	1.751	1.000	.12	.12	
	2	.249	2.653	.08	.08	
2	1	2.461	1.000	.01	.02	.01
	2	.514	2.189	.00	.26	.05
	3	.025	9.913	.99	.72	.94
3	1	3.349	1.000	.00	.01	.00
	2	.581	2.400	.00	.07	.06
	3	.045	8.585	.01	.79	.65
	4	.024	11.786	.99	.14	.89
4	1	4.130	1.000	.00	.00	.00
	2	.589	2.647	.00	.06	.05
	3	.229	4.243	.02	.03	.03
	4	.028	12.173	.11	.91	.22
	5	.023	13.305	.87	.00	.89

مقادیر ویژه (ستون **Eigenvalue**) برای تعیین تعداد ابعاد موجود در متغیرهای مستقل می‌باشند.

Model	Dimension	Eigenvalue	Condition Index
1	1	1.751	1.000
	2	.249	2.653
2	1	2.461	1.000
	2	.514	2.189
	3	.025	9.913
3	1	3.349	1.000
	2	.581	2.400
	3	.045	8.585
	4	.024	11.786
4	1	4.130	1.000
	2	.589	2.647
	3	.229	4.243
	4	.028	12.173
	5	.023	13.305
5	1	5.014	1.000
	2	.672	2.732
	3	.249	4.485
	4	.035	11.888
	5	.026	14.006
	6	.004	30.781

وقتی چندین مقدار ویژه نزدیک به صفر هستند، متغیرها شدیداً همبسته هستند و تغییرات کوچک در مقادیر داده‌ها ممکن است به تغییرات بزرگی در برآوردهای ضرائب ختم شود.



شاخص‌های شرطی (ستون **Condition Index**) ریشه‌های دوم نسبت‌های بزرگترین مقدار ویژه به هر مقدار ویژه متوالی هستند. یک شاخص شرطی بزرگتر از 15 نشان دهنده یک مسئله احتمالی است و یک شاخص بزرگتر از 30 حکایت از یک مشکل جدی با خطی بودن دارد.

**Variance Proportions** نسبت‌های واریانس برآورد در نظر گرفته شده با هر مولفه اصلی مرتبط با هر مقدار ویژه می‌باشد. هم‌خطی یک مسئله و مشکل است وقتی که مؤلفه با شاخص شرطی در ارتباط است که به طور قابل توجهی با واریانس دو یا چند متغیر مشترک است.

## متغیرهای کنار گذاشته شده رگرسیون

این جدول اطلاعات مربوط به متغیرهایی را که در هر مرحله در مدل نیستند نمایش می‌دهد. **Beta In** ضریب رگرسیون استاندارد شده است به شرطی که متغیر در هر مرحله وارد شده باشد. آماره **t** و سطح معنی‌داری (**Sig.**) آن برای بررسی فرضیه صفری به کار می‌روند که می‌گوید ضریب رگرسیون صفر است (یا اینکه هیچ رابطه خطی بین متغیر وابسته و مستقل وجود ندارد). اگر سطح معنی‌داری کوچک باشد ( $<0.05$ ) آنگاه ضریب معنی‌دار در نظر گرفته می‌شود.

Model		Beta In	t	Sig.	Partial Correlation
1	Fertility	-.119	2.536	.013	-.242
	%Urban	.112	2.980	.004	.282
	Log GDP	.121	2.630	.010	.251
	Birth/Death Rate	.025	.942	.348	.092
	Population Increase	-.002	-.050	.974	-.006
2	Fertility	-.115	2.576	.013	-.243
	Log GDP	.079	1.595	.114	.156
	Birth/Death Rate	.021	.793	.429	.078
	Population Increase	-.011	-.134	.739	-.033
3	Log GDP	.078	1.610	.109	.169
	Birth/Death Rate	.109	3.421	.001	.322
	Population Increase	.129	2.654	.009	.265
4	Log GDP	.144	3.075	.003	.294
	Population Increase	-.043	-.457	.648	-.046
5	Population Increase	-.030	-.335	.738	-.034
6	%Urban	.044	1.152	.252	.114
	Population Increase	-.011	-.121	.904	-.012

همبستگی جزئی (**Partial Correlation**) عبارت است از همبستگی هر متغیر مستقل با متغیر وابسته بعد از حذف اثر خطی متغیرهایی که قبلاً در مدل بوده‌اند.

هم خطی بودن (یا هم خطی چندگانه) وضعیت نامطلوبی است که در آن همبستگی‌های متغیرهای مستقل قوی می‌باشند.

**Tolerance** یک آماره‌ای است که مشخص می‌کند چقدر متغیر مستقل به طور خطی با دیگری وابسته است. تلرانس بخشی از واریانس متغیر است که با متغیرهای مستقل دیگر موجود در مدل در نظر گرفته نشده است. یک متغیر با تلرانس خیلی پایین اطلاعاتی از مدل را ارائه کرده، و می‌تواند باعث مشکلات محاسباتی شود.

**VIF**، یا عامل افزایش واریانس عبارتست از معکوس واریانس. با افزایش **VIF**، واریانس ضریب رگرسیون آن را برآوردی ناپایدار می‌نماید. مقادیر **VIF** بزرگ نشانه‌ای از هم خطی چندگانه می‌باشند. **Minimum Tolerance** کمترین تلرانس همه متغیرهای مستقل است که قبلاً در مدل می‌باشند به شرطی که متغیر مستقل که در معادله نیست به حساب آید.



Model		Partial Correlation	Collinearity Statistics		
			Tolerance	VIF	Minimum Tolerance
1	Fertility	-.242	.308	3.246	.308
	%Urban	.282	.470	2.127	.470
	Log GDP	.251	.322	3.106	.322
	Birth/Death Rate	.092	.988	1.012	.988
	Population Increase	-.006	.647	1.546	.647
2	Fertility	-.243	.308	3.250	.308
	Log GDP	.156	.269	3.714	.269
	Birth/Death Rate	.078	.985	1.016	.985
	Population Increase	-.033	.642	1.559	.642
3	Log GDP	.159	.269	3.714	.269
	Birth/Death Rate	.322	.565	1.771	.565
	Population Increase	.255	.253	3.947	.253
4	Log GDP	.294	.241	4.148	.241
	Population Increase	-.046	.066	15.218	.066
5	Population Increase	-.034	.066	15.250	.066
6	%Urban	.114	.370	2.700	.370
	Population Increase	-.012	.068	14.748	.068

مبحث را در مرجع کاربردی SPSS 20 (40) دنبال نمایید.



**Telegram.me/iepnu**  
 کانال تخصصی مهندسی صنایع دانشگاه پیام نور