

## ۸

## رگرسیون چند متغیره در علوم اجتماعی: آنالیز نمرات دانشجویان

نوشته: کریم منصورفر

در دهه های اخیر پیشرفت در زمینه تجزیه و تحلیل داده های آماری به وسیله کامپیوتر، آنچنان سریع بوده که پژوهشگران و محققین اعم از جامعه شناس و غیر جامعه شناس به محاسبات پر زحمت آماری فکر نمی کنند، بلکه بیشتر درباره معانی داده ها می اندیشند و به تعبیر و تفسیر داده ها می پردازند. این نوع برنامه ها در اغلب مراکز کامپیوتری وجود دارد و دیگر جایی برای نگرانی برنامه ریزی کامپیوتری باقی نمانده است. بخصوص برنامه رگرسیونی چند متغیری اکثر قریب به اتفاق در دستگاه های کامپیوتری وجود دارد. کارشناسانی که سابقه چندان در داده پردازی ندارند، استفاده از برنامه های کامپیوتری آماده را نسبتاً سریع فرا می گیرند. اصولاً کار با کامپیوتر سبب می شود که این کارشناسان اعتماد به نفس پیدا کنند و مهارت تحلیل خود را بالا ببرند و مجذوب امکانات بیکران تحلیل کامپیوتری شوند. تحلیل کامپیوتری نه فقط به تشریح مباحث کمک می کند، بلکه ما را از محاسبات پر زحمت می رهاوند و فرصت می دهد تا درباره معانی داده ها بیشتر باندیشیم و به تعبیر و تفسیر عمیق داده ها پردازیم؛ وانگهی کامپیوتر این امکان را به وجود آورده که مقادیر بسیار زیادی از داده های پیچیده، نسبتاً با

سهولت تحلیل شود. بنابراین تا آنجایی که امکان دارد باید کلیه عملیات با استفاده از کامپیوتر انجام گیرد. آنچه به تصویر کشیده شده تا حدودی هدف این تحقیق را روشن می کند، اگر این تحقیق بخوبی فهمیده شود، فراگیری روشهای چند متغیری دیگر آسانتر بوده و مهمتر از آن موارد کاربردشان در تحقیقهای مربوط به علوم اجتماعی روشنتر خواهد بود. در اینجا، اثر همزمان نمرات ۱۵ درسی که دانشجویان علوم اجتماعی قبل از انتخاب گرایش انتخاب کرده اند را بر روی معدل دو سال آخر مورد تحقیق قرار می دهیم. هدف پیدا کردن درس یا دروسی است که با معدل دو سال آخر کاملاً رابطه منطقی دارند و همچنین مشخص کردن میزان اثر آنها در پیشرفت نمرات دانشجویان. پس معدل دو سال آخر را به عنوان متغیر وابسته و ۱۵ درسی که قبل از انتخاب گرایش رشته تحصیلی، دانشجو موظف است انتخاب نماید را متغیر مستقل نامیده ایم. بدین طریق:

$y$  معدل دو سال آخر

$x_1$  ریاضیات پایه و مقدمات آمار

$x_2$  فارسی یک

$x_3$  مقدمات زبان انگلیسی

$x_4$  مقدمات جامعه شناسی

$x_5$  مقدمات انسان شناسی

$x_6$  مبانی تعاون

$x_7$  مبانی جمعیت شناسی

$x_8$  کلیات حقوق

$x_9$  معارف اسلامی

$x_{10}$  جامعه شناسی روستایی

$x_{11}$  اصول سازمان و مدیریت

$x_{12}$  اصول علم اقتصاد

$x_{13}$  آمار در علوم اجتماعی

$x_{14}$  مقدمات روانشناسی

$x_{15}$  جغرافیای انسانی ایران

داده‌ها از کارنامه ۳۲ دانشجوی فارغ التحصیل که به طور تصادفی انتخاب شده‌اند، جمع‌آوری گردیده و کوچکترین تغییری انجام نگرفته است.

$X =$

154	100	180	125	170	140	160	170	162	160	165	170	130	100	165	128
154	100	130	135	150	170	160	120	164	190	145	120	170	115	175	140
158	050	190	107	150	160	120	150	148	160	170	160	165	100	183	190
156	180	160	185	170	190	130	170	180	170	140	160	165	155	189	130
150	125	130	180	130	160	200	142	092	140	190	150	140	130	150	125
157	185	170	166	155	190	170	160	166	185	140	130	140	140	200	145
173	150	200	180	200	200	175	170	182	195	170	165	170	165	182	185
166	140	170	136	150	170	160	145	172	170	155	155	160	114	190	135
140	170	140	170	140	160	200	170	148	200	125	125	170	080	190	105
147	120	145	124	100	140	120	150	152	165	155	135	120	110	160	150
163	180	180	170	140	155	150	175	125	180	150	160	180	128	185	100
152	140	165	176	140	170	130	150	185	170	110	140	150	140	180	170
156	155	145	146	180	200	200	175	176	170	185	135	130	150	178	180
167	175	160	165	170	190	170	152	188	190	175	140	170	130	190	140
158	170	140	175	150	140	190	150	175	150	155	125	150	134	188	150
158	160	175	150	170	140	150	160	168	135	135	150	130	114	190	100
159	160	150	078	190	140	135	170	152	180	160	115	155	175	165	115
137	050	140	135	100	140	130	130	170	150	195	125	130	150	112	160
141	020	170	083	100	170	120	105	152	145	160	110	170	080	133	150
147	045	160	133	140	130	170	140	125	140	175	130	160	120	115	160
155	120	140	135	155	120	155	172	115	140	180	150	155	135	174	165
140	000	140	158	150	130	130	110	110	150	155	140	128	130	156	150
142	065	170	100	200	190	195	140	170	170	195	140	140	125	184	170
154	120	160	156	195	140	185	175	190	190	140	150	170	120	170	165
158	100	160	100	190	160	155	125	150	155	155	145	175	125	180	145
157	140	130	150	175	185	185	170	200	150	150	145	175	132	158	145
177	180	165	120	155	170	200	175	195	170	170	165	185	140	200	190
142	110	170	101	150	170	190	160	150	170	170	145	175	180	182	122
150	120	120	157	128	155	100	102	125	165	130	115	150	125	128	062
146	170	150	140	140	190	200	180	068	190	170	130	175	120	190	170
159	120	170	178	100	180	170	140	175	160	062	125	170	114	162	140
160	060	150	086	165	110	130	160	170	180	150	155	145	150	192	175

اولین جدولی که کامپیوتر در این رابطه در اختیار گذارده، جدول میانگین‌ها و انحراف معیارهای متغیرها می‌باشد. به طوری که ملاحظه می‌شود میانگین نمرات آمار در علوم اجتماعی از تمامی دروس به جز ریاضیات پایه و مقدمات آمار کمتر است. از طرفی انحراف معیار نمرات ریاضیات پایه و مقدمات آمار از کلیه دروس بیشتر است. این مطلب نشانگر این حقیقت است که هماهنگی و یکنواختی بین نمرات ریاضیات پایه، خیلی کمتر است و تشدد نمرات این درس نسبت به سایر دروس بیشتر است ( $S_1 = 5,074$ ). به سخن دیگر سطح دانش ریاضیات پایه، اکثریت قریب به اتفاق دانشجویان از حد نصاب پایینتر بوده و دلیل کمتر بودن میانگین نمرات ریاضیات پایه نسبت به سایر دروس به همین علت بوده است. متأسفانه این مسئله سال بسال بارزتر می‌گردد. انگیزه عمده گذاردن درس ریاضیات پیش دانشگاهی در اکثر رشته‌های مختلف دانشگاه همین بوده است. بعد از نمرات ریاضیات پایه بیشتر تشدد نمرات در درس مقدمات زبان انگلیسی است، و از لحاظ معدل نیز نسبت به سایر دروس به جز آمار در علوم اجتماعی و ریاضیات پایه سطح پایینتری قرار گرفته است با وجود اینکه اغلب دانشجویان زبان پیش دانشگاهی را نیز گذرانده‌اند، با این حال اکثر آنها از لحاظ دانش زبان خارجی در سطح پایینتری قرار دارند. در واقع مطالب بیان شده در مورد ریاضیات پایه درباره زبان خارجه نیز صادق است. وانگهی در درجه اول میانگین نمرات درس مقدمات روانشناسی (۱۷,۱۷۵) و در درجه دوم میانگین نمرات درس معارف اسلامی (۱۶,۶۷۲) از همه دروس بیشتر است. طبیعی است میزان پراکندگی نمرات این دو درس، نیز نسبت به سایر دروس خیلی کمتر است. این مشخص کننده‌های آماری، بیانگر این مطلب است که تقریباً ۹۹,۹ درصد دانشجویان در این دو درس نمره الف گرفته‌اند. نامنتقی بودن نمرات این دو درس کاملاً روشن است. بنابراین نمرات این دو درس خارج از حد نرمال بوده و هرگز با منحنی نرمال تطابق نمی‌کنند و احتیاج به آزمون این نمرات نیست.

با مقایسه میانگین‌ها و انحراف معیار کلیه دروس مورد مطالعه، این مسئله

کاملاً آشکار است که میزان یکنواختی نمرات درس اصول سازمان و مدیریت، از تمامی دروس بیشتر بوده ( $S_{11} = 1,612$ )، و از لحاظ معدل نیز تا حدودی در سطح معقولی قرار گرفته است و با استفاده از آزمون  $\chi^2$  معلوم گردیده که نمرات این درس کاملاً بر روی منحنی نرمال منطبق است.

Listwise Deletion of Missing Data

	Mean	Std Dev	Label
Y	15.416	.950	
X1	12.125	5.074	
X2	15.703	1.879	
X3	14.063	3.122	
X4	15.281	2.851	
X5	16.109	2.412	
X6	16.047	2.933	
X7	15.197	2.186	
X8	15.625	3.048	
X9	16.672	1.799	
X10	15.569	2.631	
X11	14.079	1.612	
X12	15.619	1.815	
X13	12.894	2.304	
X14	17.175	2.309	
X15	14.553	2.899	

\*\*\* MULTIPLE REGRESSION \*\*\*  
 N of Cases = 32

Correlation:

	Y	X1	X2	X3	X4	X5	X6	X7
Y	1.000	.538	.381	.158	.412	.213	.133	.411
X1	.538	1.000	.035	.483	.240	.415	.404	.649
X2	.381	.035	1.000	-.096	.258	.253	.006	.276
X3	.158	.483	-.096	1.000	-.098	.287	.218	.149
X4	.412	.240	.258	-.098	1.000	.177	.343	.449
X5	.213	.415	.253	.287	.177	1.000	.393	.139
X6	.133	.404	.006	.218	.343	.393	1.000	.488
X7	.411	.649	.276	.149	.449	.139	.488	1.000
X8	.435	.205	.265	.037	.326	.259	.052	.178
X9	.243	.433	.156	.114	.300	.401	.220	.371
X10	-.110	-.212	-.040	-.369	.251	-.033	.206	.084
X11	.514	.150	.511	.102	.443	-.006	.139	.507
X12	.261	.264	.213	.000	.118	.344	.274	.189
X13	.244	.241	-.014	-.002	.367	.101	.073	.292
X14	.524	.591	.357	.100	.501	.319	.364	.570
X15	.171	-.241	.265	-.196	.202	.131	.220	.212

	X8	X9	X10	X11	X12	X13	X14	X15
Y	.435	.243	-.110	.514	.361	.244	.524	.171
X1	.205	.433	-.212	.150	.264	.241	.591	-.241
X2	.265	.156	-.040	.511	.213	-.014	.357	.265
X3	.037	.114	-.369	.102	.000	-.002	.100	-.196
X4	.326	.300	.251	.443	.118	.367	.501	.202
X5	.259	.401	-.033	-.006	.344	.101	.319	.131
X6	.052	.220	.206	.109	.274	.073	.364	.220
X7	.178	.371	.084	.507	.189	.292	.370	.212
X8	1.000	.196	-.225	.177	.032	.167	.256	.227
X9	.196	1.000	-.153	.015	.357	.136	.508	.047
X10	-.225	-.153	1.000	.198	-.184	.229	.126	.334
X11	.177	.015	.198	1.000	.088	.150	.449	.261
X12	.082	.357	-.184	.088	1.000	-.016	.248	.037
X13	.167	.136	.229	.150	-.016	1.000	.132	.125
X14	.256	.508	-.126	.449	.248	.132	1.000	.168
X15	.227	.047	.334	.261	.037	.125	.168	1.000

موقعیت آرمانی رگرسیون چند متغیری ایجاب می کند که همبستگیهای میان بعضی از متغیرهای مستقل ضعیف و بین بعضی از متغیرهای مستقل و متغیر وابسته قوی باشند. در این جا، آنچه ما داریم، مخلوطی است از ضرایب همبستگی، تا زمانی که تحلیل انجام نشده است به هیچ وجه نمی توان آنها را از هم جدا کرد. جدول بعدی این مطلب را روشن می سازد.

جدول دوم نشانگر مطالب زیادی است، ستون اول میزان همبستگی معدل دو سال آخر دانشجویان را با دروسی که مورد مطالعه قرار گرفته، نشان می دهد. میزان و درجه همبستگی معدل دو سال آخر نسبت به درس ریاضیات پایه و مقدمات آمار از تمامی دروس مورد مطالعه بیشتر است ( $\pi_{y_1} = 0,528$ ) یعنی هر دانشجویی که نمره ریاضیات پایه و مقدمات آمار، آن بهتر و یابدتر بوده با 99 درصد، اطمینان معدل دو سال آخر نیز بهتر و یا بدتر بوده است. این رابطه نزدیک را هیچ کدام از دروس مورد مطالعه دیگر ندارد. از طرفی نمره ریاضیات پایه و مقدمات آمار با 99 درصد اطمینان ( $\pi_{(x_1, y_1)} = 44.87\%$ ) همبستگی مستقیمی با دروس مبانی جمعیت شناسی ( $\pi_{(x_1, y_1)} = 0,649$ ) و مقدمات روانشناسی ( $\pi_{(x_1, y_1)} = 0,591$ ) و زبان خارجه ( $\pi_{(x_1, y_1)} = 0,483$ ) دارد. یک مسئله اساسی که باید اینجا گفته شود، این است که نمره ریاضیات پایه، تقریباً با نصف نمره دروس مورد مطالعه رابطه مستقیم و مثبتی دارا است که هیچ کدام از دروس دیگر نشان دهنده چنین رابطه ای نیست. یک مسئله دیگر که این جدول نشان می دهد ناهمبسته بودن ریاضیات پایه و مقدمات آمار با آمار در علوم اجتماعی است که بر خلاف انتظار نمرات این دو درس، ربط آنچنانی نسبت به هم نشان نمی دهند. شاید علتش محتویات قبلی درس آمار در علوم اجتماعی باشد. زیرا قبلاً درس آمار در علوم اجتماعی را فقط مبحث احتمالات تشکیل می داد. همچنین ریاضیات پایه و مقدمات آمار، با نمره درس فارسی کوچکترین رابطه ای را نشان نمی دهد بدین معنی که افزایش یا کاهش نمره درس ریاضیات پایه، ربطی به نمره درس فارسی ندارد، ولی درس فارسی با نمره درس اصول سازمانهای مدیریت،

بیشترین همبستگی مستقل را داراست. به عبارتی هر چقدر ادبیات فارسی دانشجوی بهتر بوده، به همان نسبت درس اصول سازمانهای مدیریت را بهتر یاد گرفته و سئوالات امتحانی را گویاتر نوشته است. از آنجایی که درس جامعه‌شناسی یکی از دروس بسیار مهم و اساسی این دانشجویان می‌باشد، لذا میزان رابطه نمره مقدمات جامعه‌شناسی با ۹۹ درصد اطمینان با نمرات درس مبانی جمعیت‌شناسی، اصول سازمان و مدیریت، مقدمات روانشناسی و با ۹۵ درصد اطمینان نیز با نمره درس آمار در علوم اجتماعی همبستگی دارد. یک نکته مهم که در این جدول به چشم می‌خورد، همبستگی معکوس نمره جامعه‌شناسی روستایی با اکثریت قریب به اتفاق نمرات دروس دیگر است. این همبستگی معکوس را هیچ کدام از دروس دیگر دارا نیست. چنین به نظر می‌رسد که اکثریت دانشجویان مورد مطالعه از شهرهای کوچک یا خیلی کوچک باشند. بدین دلیل دانشجویان شناخت بیشتری با جامعه روستایی دارند. حال به همبستگی سایر دروس نسبت به هم می‌پردازیم.

مقدمات انسان‌شناسی با ۹۹ درصد اطمینان با معارف اسلامی و با ۹۵ درصد اطمینان با مبانی تعاون همبستگی دارد این همبستگیها منطقی به نظر می‌آید. زیرا یکی از مهمترین مسائلی که در جهان بینی هر مکتب وجود دارد، شناخت انسان از دیدگاه آن مکتب است. شاید همبستگی نمرات انسان‌شناسی با نمرات معارف اسلامی همین امر باشد، وانگهی از آنجایی که همه احکام اجتماعی از جمله مبانی تعاون به شناخت انسان مربوط می‌شود، باید نمرات مقدمات انسان‌شناسی از نمرات مبانی تعاون پیروی نماید.

به طور کلی ملاحظه می‌شود به جز ریاضیات پایه که با اکثر دروس همبستگی دارد، بقیه دروس کم و بیش بسته به اشتراکی که دروس نسبت بهم دارند، رابطه مستقیمی را نشان می‌دهند، ولی تنها نمرات دو درس است که با هیچ کدام از دروس مورد مطالعه اصولاً همبستگی ندارند. این دروس عبارتند از کلیات حقوق و جغرافیای انسانی ایران. چرا نمرات این دو درس با سایر دروس



همبستگی ندارد؟ دلیل این امر شاید این باشد که محتوای این دو درس با سایر دروس هماهنگی نداشته و شاید هم یک مقوله دیگری باشد.

حال به اصل اساسی مطلب که هدف این تحقیق است، می پردازیم. از آنجایی که رگرسیون چند متغیری و زیربنای منطقی آن، شالوده اغلب روشهای چند متغیری بوده و در نتیجه نظام عمومی تجزیه و تحلیل داده های آماری است، ضروری است به طور خلاصه سخنی کوتاه درباره رگرسیون و همبستگی داشته باشیم.

در مطالعه ارتباط بین صفات، باید بین مفاهیم همبستگی صفات با روابط علت و معلولی تفاوت گذاشت. وقتی بین صفات ارتباط باشد، لازم نیست حتماً هر یک از آنها علت دیگری باشد. از طرفی تحلیل رگرسیون ساده، نشانگر ارتباط خطی متغیرها با یکدیگر است. در این تحلیل، یک تابع کامل یعنی معادله رگرسیون برآورده می شود که مقادیر متغیر وابسته را بر حسب تغییرات صفات دیگر بیان می کند.

اما در تحلیل همبستگی که میزان و درجه ارتباط خطی متغیرها را به صورت عدد شاخص نشان می دهد حد و حدود تغییرات و روند حرکت متغیرها به گونه ای آنی تصویر می شود. در واقع تحلیل همبستگی این نکته را روشن می سازد که متغیرها تا چه حد همراه هم حرکت می کنند. اگر چه همبستگی تکنیکی است که قدرت کمتری از رگرسیون دارد. این دو، چنان ارتباط نزدیکی با هم دارند که همبستگی غالباً کمک مفیدی در تعبیر و تفسیر رگرسیون است، در واقع دلیل عمده بررسی همبستگی نیز همین بوده است. این مفهوم در تعبیر و تفسیر ضریب تعیین ( $\pi^2$ ) کاملاً روشن است.

حال صورت کلی معادله رگرسیون چند متغیری را در نظر می گیریم و فهرست وار طرز محاسبه را گویا شده و آنگاه آن را در مورد این تحقیق انجام می دهیم، بنابراین معادله رگرسیون عبارت است از:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_{15}x_{15} \quad (1)$$

مقدار  $b$  و ضرایب  $b$  در این معادله باید محاسبه شود تا بتوان در صورت لزوم از طریق متغیرهای مستقل ( $x$  ها یعنی نمرات دانشجویان) متغیر وابسته ( $y$  معدل دو سال آخر) مربوط به هر فرد را پیش بینی نمود.

این مقادیر با استفاده از روش حداقل مربعات حل می شود (برای اطلاع بیشتر به کتاب روشهای آماری مراجعه شود) در اینجا رابطه بین متغیر وابسته ( $y$ ) و متغیرهای مستقل ( $x_1, x_2, x_3, \dots, x_{15}$ ) تابع خطی در نظر گرفته شده است. همان طوری که قبلاً اشاره شد در این تحقیق فارغ التحصیلان یک ترم را انتخاب کرده و از بین آنها دانشجویانی که تقریباً این ۱۵ درس را در ترمهای اول تا چهارم انتخاب نموده اند، در نظر گرفته ایم، مسئله ای که در اینجا مشکل ایجاد می کرد، بعضی دروس بود که انتخاب آن اجباری بود. ولی عده ای از دانشجویان به علل ضعف در آن مورد، آن دروس را حذف کرده و بجای آن دروس دیگری انتخاب کرده اند. به همین دلیل بود که کلیه فارغ التحصیلان یک ترم، در نظر گرفته نشده، از طرفی بعضی از دروسی که دانشجو باید در عرض چهار ترم اول انتخاب نماید، اما به عللی انتخاب نکرده، نیز در این محاسبات منظور نشده است.

چون اکثر خوانندگان این مقالات محققین و به خصوص دانشجویان می باشند، لذا عملیات مربوط به طور خلاصه بیان می گردد.

برای اینکه در فرمول (۱) مقادیر  $b$ ها محاسبه گردد، باید از فرمول

$$b = SP^{-1}.g \quad (2)$$

استفاده شود. برای این کار باید مراحل زیر را طی کرد:

- ۱- ماتریس  $x$  را تشکیل داد در این ماتریس، ستون اول مقادیر  $y$ ها و ستون دوم مقادیر  $x_1$  ها و ستون سوم مقادیر  $x_2$ ها و ... ستون شانزدهم مقادیر  $x_{15}$ ها می باشد. سپس ترانها ده ماتریس  $x$  را پیدا کرده آن را  $x'$  می نامیم.
- ۲- ماتریس  $x'x$  را محاسبه می کنیم.
- ۳- ماتریس  $CF$  را تشکیل می دهیم.
- ۴- ماتریس  $CF$  را از ماتریس  $x'x$  کم می کنیم.



۵- با استفاده از ماتریس  $x'x$ -CF ماتریس های  $g$ ،  $SP$  را به دست می آوریم،  
 آنگاه معکوس ماتریس  $SP$  را محاسبه می کنیم و آن را  $SP^{-1}$  مقادیر می نامیم.

$$g = \begin{pmatrix} SP_{1y} \\ SP_{2y} \\ SP_{3y} \\ SP_{4y} \\ SP_{5y} \\ SP_{6y} \\ SP_{7y} \\ SP_{8y} \\ SP_{9y} \\ SP_{10y} \\ SP_{11y} \\ SP_{12y} \\ SP_{13y} \\ SP_{14y} \\ SP_{15y} \end{pmatrix} = \begin{pmatrix} 80,39 \\ 21,05 \\ 14,56 \\ 34,56 \\ 15,15 \\ 11,48 \\ 26,47 \\ 39 \\ 12,86 \\ -8,48 \\ 24,41 \\ 19,27 \\ 16,58 \\ 35,61 \\ 14,55 \end{pmatrix}$$

۶- با استفاده از ماتریس  $g$  و ماتریس  $SP^{-1}$  مقادیر  $b$  ها را از فرمول (۲) محاسبه می کنیم. همانطور که قبلاً اشاره شد، کلیه محاسبات را می توان براحتی با کامپیوتر انجام دارد، زیرا برنامه های ویژه این محاسبات در اغلب مراکز کامپیوتری وجود دارد. در اینجا فقط ماتریسهای  $x$  و  $x'x$ -CF و  $g$  آمده است زیر عملیات به طور مستقیم با کامپیوتر انجام گرفته است.



\*\*\* MULTIPLE REGRESSION \*\*\*

Equation Number 1 Dependent Variable.. Y

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
X15	.16440	.05981	.50177	2.749	.0143
X12	.08649	.07718	.16527	1.121	.2790
Y13	7.450519E-03	.05656	.01807	.132	.8968
X3	-.07151	.05919	-.23510	-1.209	.2441
X8	.02200	.04741	.07062	.464	.6488
X11	.32586	.12667	.55309	2.572	.0205
X9	.01359	.08523	.02574	.159	.8753
X6	-.01584	.05692	-.04891	.278	.7844
X5	-.11296	.06934	-.28690	-1.629	.1228
X2	.08061	.08227	.15946	.980	.3418
X4	.07435	.05715	.22320	1.301	.2117
X14	-.11967	.09753	-.29095	-1.227	.2376
X7	-.33057	.11586	-.76071	-2.853	.0115
X10	-.06055	.06890	-.16775	-.879	.3925
X1	.24827	.06087	1.32628	4.079	.0009

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
(Constant)	12.10727	2.41707		5.009	.0001

End Block Number 1 All requested variables entered.

در اینجا ضریب همبستگی چندگانه (چند متغیر) برابر است با  $R = 0,88623$ . این مقدار ضریب نشان دهنده رابطه شدید بین معدل دو سال آخر (y)، با نمرات ۱۵ درس انتخاب شده است، البته این ضریب همیشه مثبت است و بین صفر و یک می باشد  $0 \leq R \leq 1$  هر چقدر به یک نزدیکتر گردد، میزان وابستگی متغیر وابسته به متغیرهای مستقل، بیشتر است.

$$R = \sqrt{\frac{SS_R}{SS_T}} = 0,88623$$

همان طوری که اشاره شد:

$$R^2 = \frac{P(1-R^2)}{N-P-1} = 0,58423$$

و R تعدیل یافته نیز برابر با

(که در آن  $P=15$  تعداد متغیرهای مستقل می باشد)

و میزان خطای معیار یا خطاهای همبستگی برابر با  $S_E = 0,61239$  می باشد. این شاخص نشان دهنده میزان قدرت پیش بینی معادله رگرسیون چند متغیری می باشد.

ضریب تعیین نیز در اینجا برابر با  $R^2 = 0,7854$  است این عدد بیانگر این حقیقت است که  $78,54$  درصد تغییرات معدل دو سال آخر به نمرات دو سال اول مربوط است و  $21,45$  درصد مربوط به عوامل دیگر است.

در اینجا آنالیز واریانس نیز انجام گردیده و مقدار F یعنی نسبت متوسط واریانس رگرسیون به متوسط واریانس باقیمانده محاسبه گردیده برابر با  $F = 3,90407$  و معنی دار بودن F برابر با  $0,0051$  یعنی با  $99,5$  درصد اطمینان بین نمرات اختلاف معنی دار وجود دارد. به عبارتی نمرات یکسان نیستند. البته می توان مقدار F را از فرمول

$$F = \frac{\frac{R^2}{K}}{\frac{1-R^2}{N-K-1}}$$

نیز محاسبه کرد که در آن  $N=32$  حجم و  $K=15$  تعداد متغیرهای مستقل است در نتیجه  $F=3,90425$  تقریباً با همان مقدار قبلی برابر است و اختلاف بسیار ناچیز نیز ناشی از گرد کردن ارقام می باشد.

لازم است که اهمیت نسبی نمرات دروس، در پیش بینی معدل دو سال آخر شناخته شود. لذا باید  $b$ ها و ارسی شوند. البته کلیه  $b$ ها (ضرایب مجهولات) توسط کامپیوتر محاسبه گردیده است. بنا براین معادله کلی خط رگرسیون عبارت است از:

$$y = 0,24827x_1 + 0,08061x_2 - 0,07151x_3 + 0,07435x_4 - 0,11295x_5 \\ - 0,01584x_6 - 0,33057x_7 + 0,02200x_8 + 0,01359x_9 \\ - 0,06055x_{10} + 0,32586x_{11} + 0,08649x_{12} + 0,00745x_{13} \\ - 0,11967x_{14} + 0,16440x_{15} + 12,10727$$

با توجه به  $b$ های محاسبه شده (علامت  $b$ ها) ملاحظه می شود که اثر نمرات دروس ریاضیات پایه ( $x_1$ )، فارسی ( $x_2$ )، مقدمات جامعه شناسی  $x_4$ ، کلیات حقوق ( $x_8$ )، معارف اسلامی ( $x_9$ )، اصول سا زمان و مدیریت ( $x_{11}$ )، اصول علم اقتصاد ( $x_{12}$ )، آمار در علوم اجتماعی ( $x_{13}$ ) و جغرافیای انسانی در ایران ( $x_{15}$ ) بر روی معدل دو سال آخر به طور مستقیم بوده در صورتی که نمرات سایر دروس چنین حالتی را ندارند. می توان با قرار دادن نمرات یک دانشجو قبل از انتخاب رشته تحصیلی معدل دو سال آخر را پیشگویی کرد. مثلاً در مورد دانشجوی اولی  $\hat{y} = 15,39964 + 15,4$  همان مقدار  $y_1$  به دست آمده، بدینسان می توان هر دانشجویی را که جزو نمونه نیست، معدل دو سال آخرش را برآورد نمود. در واقع



نتایج تحلیل رگرسیون چند متغیری بخوبی برازنده پیش بینی است. معادله پیدا شده در بالا نسبت به دستگاه محورهای مختصات عادی است. می توان آن را به معادله در دستگاه استاندارد تبدیل نمود. زیرا بیان معادله همبستگی بر حسب صفات استاندارد یا به عبارت دیگر بیان معادله همبستگی در دستگاه استاندارد نه تنها ضرایب آن معادله را با همدیگر قابل مقایسه می نماید، بلکه فرمولهای مشخص کننده آماری و خود معادلات را نیز ساده تر می کند. صورت کلی معادله همبستگی در دستگاه استاندارد عبارت است از:

$$Uy = Z = \beta_1 u_1 + \beta_2 u_2 + \dots + \beta_k U_k$$

برای تحقیق مورد مطالعه مقادیر  $\beta$  با استفاده از کامپیوتر محاسبه گردیده:

$$Uy = 1,32628u_1 + 0,15946u_2 - 0,23510u_3 + 0,22320u_4 - 0,28690u_5 - 0,04891u_6 - 0,76071u_7 + 0,07062u_8 + 0,02574u_9 - 0,1677u_{10} + 0,55309u_{11} + 0,16527u_{12} + 0,01807u_{13} - 0,29095u_{14} + 0,50177u_{15}$$

در این معادله، تمامی ضرایب ( $\beta$ ) ها با هم قابل مقایسه هستند، و در آن

$$U = \frac{x_i - \bar{x}}{S_x} \quad \beta = b \frac{S_x}{S_y}$$

از طرفی  $\beta$  ها تأثیر هر درس را به طور جدا از تأثیر دروس دیگر بر معدل دو سال آخر نشان می دهد. بنابراین اندازه ای است خالص و استاندارد. لازم است گفته شود که  $\beta$  های مثبت اثر مستقیم و  $\beta$  های منفی اثر معکوس بر روی معدل دو سال آخر را دارا می باشند. حال می توان تک تک نمرات را بر روی معدل دو سال آخر بررسی نمود. ملاحظه می شود که از بین این ضرایب،  $\beta$  ی مربوط به ریاضیات پایه و مقدمات آمار از همه بزرگتر است. به سخن دیگر اثر این درس بر روی معدل دو سال آخر نسبت به سایر دروس چند برابر می باشد. در مرحله دوم نیز اثر مثبت نمره درس سازمانهای مدیریت بر معدل دو سال آخر بیش از سایر دروس است.

بزرگترین اثرات منفی نمرات دروسی که بر معدل دو سال آخر تأثیر می گذارند، به مبانی جمعیت شناسی، روانشناسی اجتماعی و مقدمات انسان شناسی تعلق دارد. اینکه چرا نمرات این دروس بر معدل دو سال آخر اثری منفی دارند، باید ریشه یابی گردد. باید متذکر شد که داده های آماری صد درصد واقعیات را نشان نمی دهند، بلکه تصویری از آن را دربر دارند. چه بسا ممکن است این تصویر دارای خطا باشد. لازم به تذکر است که اگر تحلیل رگرسیون شامل یک متغیر مستقل بود، مقدار  $\beta$  با ضریب همبستگی ( $\pi$ ) رابطه نزدیکی داشت

$$(\hat{\beta} = r \frac{S_y}{S_x})$$

در ستون آخر (صفحه ۲۴۴) ملاحظه می شود که تمامی  $b$ ها با استفاده از آزمون  $t$  استودنت آزمون شده اند. نمراتی که با معدل دو سال آخر تحصیلی ( $y$ ) با ۹۵ درصد اطمینان معنی دار بودند عبارتند از: ریاضیات پایه ( $x_1$ )، مبانی جمعیت شناسی ( $x_7$ )، اصول سازمانهای مدیریت ( $x_{11}$ ) و جغرافیای انسانی در ایران ( $x_{15}$ )، بدین دلیل برای نمرات این دروس جداگانه رگرسیون چند متغیری انجام گردیده است. در نتیجه، بهترین برآزش بین معدل دو سال آخر و ترکیب خطی نمرات دروس از لحاظ آماری چنین شده است:

$$y = 0,16155x_1 + 0,32393x_{11} - 0,21742x_7 + 0,11182x_{15} + 10,57314$$

برای این معادله رگرسیون نیز آزمون آنالیز واریانس محاسبه گردیده، و مقدار  $F$  برابر با  $F_c = 10,10854$  و معنی دار بودن  $F$  نیز عبارت است ( $F = 0,00000$  معنی دار) این عدد را نشان می دهد که به طور حتم معدل دو سال آخر با این چهار درس رابطه خطی مستقیمی را داراست. طبیعی است احتمال اینکه نسبت  $F_c$  به این بزرگی به طور تصادفی به دست آید، فوق العاده کم است. رابطه بین معدل دو سال آخر با دروس نامبرده نمی تواند تصادفی پیش آمده باشد.

\*\*\*\*\* MULTIPLE REGRESSION \*\*\*\*\*

Listwise Deletion of Missing Data

	Mean	Std Dev	Label
Y	15.416	.950	
X1	12.125	5.074	
X7	15.197	2.186	
X11	14.078	1.612	
X15	14.553	2.899	

N of Cases = 32

Correlations:

	Y	X1	X7	X11	X15
Y	1.000	.538	.411	.514	.171
X1	.538	1.000	.649	.150	-.241
X7	.411	.649	1.000	.507	.212
X11	.514	.150	.507	1.000	.261
X15	.171	-.241	.212	.261	1.000

\*\*\*\*\* MULTIPLE REGRESSION \*\*\*\*\*

Equation Number 1 Dependent Variable.. Y

Variable(s) Entered on Step Number  
4.. X15

Multiple R .77434  
 R Square .59961  
 Adjusted R Square .54029  
 Standard Error .64394

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	4	16.76639	4.19160
Residual	27	11.19579	.41466

F = 10.10854      Signif F = .0000

\*\*\*\*\* MULTIPLE REGRESSION \*\*\*\*\*

Equation Number 1      Dependent Variable.. Y

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
X1	.16155	.03567	.86304	4.529	.0001
X11	.32393	.06669	.54981	3.737	.0009
X7	-.21742	.09124	-.50032	-2.363	.0245
X15	.11182	.04753	.34130	2.352	.0262
(Constant)	10.57314	1.11178		9.510	.0000

End Block Number 1      All requested variables entered.

انحراف معیار این چهار درس عبارت است از:

$S_1 = 5,074$  (ریاضیات پایه و مقدمات آمار)،  $S_7 = 2,186$  (مبانی جمعیت

شناسی)

$S_{11} = 1,612$  (اصول سازمانهای مدیریت)،  $S_{15} = 2,899$  (جغرافیای انسانی

در ایران)

ملاحظه می شود که اختلاف نمرات درس اصول سازمانهای مدیریت از همه کمتر و ریاضیات پایه از همه بیشتر است. در واقع یک نواختی در نمرات اصول سازمانهای مدیریت از همه دروس بیشتر است و ضرایب همبستگی نمرات این چهار درس نسبت به معدل دو سال آخر در مرحله اول ریاضیات پایه  $R = 0,538$  و در مرحله دوم، اصول سازمانهای مدیریت  $R_{11} = 0,514$ ، و در مرحله سوم مبانی جمعیت شناسی و بالاخره در مرحله آخر جغرافیای انسانی در ایران است. نکته مهمی که در این جدول به چشم می خورد همبستگی شدید نمرات ریاضیات پایه با مبانی جمعیت شناسی است ( $\pi = 0,649$ ).

ضریب همبستگی چند متغیری  $R = 0,77434$  را می توان مانند ضریب همبستگی یک متغیری تفسیر کرد ( $0 \leq R \leq 1$ ) و مقدار ضریب تعیین  $R^2 = 0,5996$  می باشد. معنای این کمیت نیز بیانگر این است که تقریباً ۶۰ درصد واریانس نمرات معدل دو سال آخر دانشجویان مورد مطالعه توسط یک ترکیب خطی از نمرات ریاضیات پایه، اصول سازمانهای مدیریت، مبانی جمعیت شناسی و جغرافیای انسانی در ایران ناشی می شود. همان طوری که قبلاً گفته شد، در تحلیل با ۱۵ متغیر مستقل تقریباً ۷۸ درصد تغییرات معدل دو سال آخر ( $y$ ) با ۱۵ درس یاد شده مربوط بود. یعنی در واقع اضافه کردن ۱۱ درس قدرت پیش بینی را فقط ۱۸ درصد بالا می برد.

معادله استاندارد شده با چهار درس یاد شده عبارت است از:

$$Uy = 0,863.4U_1 + 0,54981U_{11} - 0,50032U_7 + 0,34130 U_{15}$$

از آنجایی که ضرایب  $Z$  ها ( $\beta$ ها) با ضریب همبستگی رابطه نزدیکی دارند،

لذا  $\beta$ های محاسبه شده نشانگر این حقیقت اند که دروس ریاضیات پایه ( $x_1$ ) و مبانی جمعیت شناسی ( $x_7$ ) به ترتیب سهم بیشتر و اصلی تری در تبیین معدل دو سال آخر ( $y$ ) را دارا می باشند. ولی چون علامت  $\beta_1$  مثبت و  $\beta_7$  منفی است بنابراین رابطه بین معدل دو سال آخر با ریاضیات پایه به طور مستقیم و با مبانی جمعیت شناسی به طور معکوس است. در واقع ریاضیات پایه بزرگترین سهم را داراست. یعنی هر دانشجویی که ریاضیات آن بهتر بوده، موفقیتش در آخر سال نسبت به سایر دانشجویان بیشتر بوده است و در مرحله دوم نیز، درس اصول سازمانهای مدیریت و در مرحله سوم، جغرافیای انسان در ایران بوده است؛ و این بهترین دلیلی است که دانشگاههای بزرگ جهان در دهه های اخیر درس ریاضیات را در رشته های مختلف دوبرابر کرده اند: از طرفی دانشجویانی که ریاضیات آنها تا حدودی خوب است، موفقیت آنها بیش از دانشجویانی است که نمره ریاضی آنها در حد پایین تری قرار دارد. بنابراین در کنکور سراسری باید دانشجویانی برای این رشته ها انتخاب شوند که دیپلم ریاضی و یا تجربی داشته باشند و نمره ریاضی آنها از حد نصاب پایینتر نباشد. وانگهی در این چند دهه اخیر در اکثر کشورهای پیشرفته جهان کلیه تحقیقات و بررسیها اعم از اقتصادی و اجتماعی با استفاده از کامپیوتر انجام می گیرد. دانشجوی علوم اجتماعی باید با آن قسمت از مباحث ریاضی که به هر شکلی در تحقیقات علوم اجتماعی جنبه اساسی دارد آشنایی بیشتری داشته باشد، و این کار وقتی امکان پذیر است که دانشجویان شرایط فوق را داشته باشند.