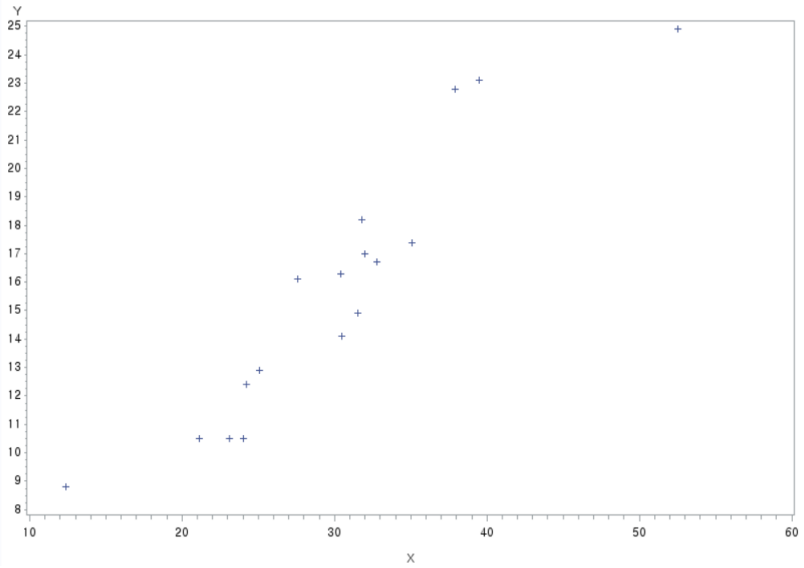
회귀분석론 과제3 오유진(2015325)

7.4

7.4.1.



7.4.2

Z = x-x̄ = x-30.1이며 z, z2, z3, z4는 각각 Y에대한 1차,2차,3차,4차 다항회귀모형을 만드는데 필요한 새로운 설명변수라고 두었다.

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read	17
Number of Observations Used	17

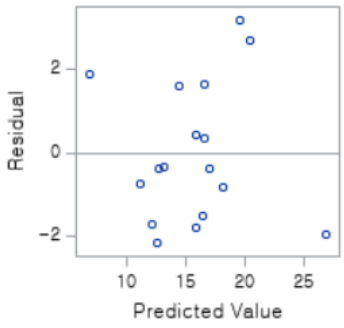
Analysis of Variance

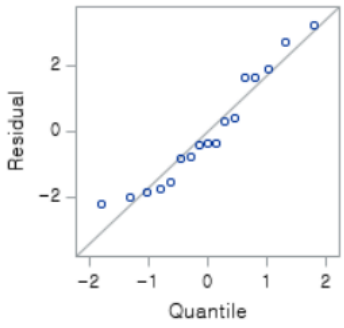
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	307.25750	307.25750	101.16	<.0001
Error	15	45.56015	3.03734		
Corrected Total	16	352.81765			

Root MSE	1.74280	R-Square	0.8709
Dependent Mean	15.71176	Adj R-Sq	0.8623
Coeff Var	11.09231		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	15.71762	0.42269	37.18	<.0001
z	1	0.49808	0.04952	10.06	<.0001





p값이 모두 유의수준 0.05보다 작으므로 y와 z는 상당히 유의적이다라는 것을 알 수 있다. 잔차산점도는 약간 이분산성을 보인다고 할 수 있고 정규확률그림을 보면 완전한 직선모양이라고 보

기는 어려우므로 개선할 필요가 있다.

The REG Procedure
Model: MODEL2
Dependent Variable: Y

Number of Observations Read	17
Number of Observations Used	17

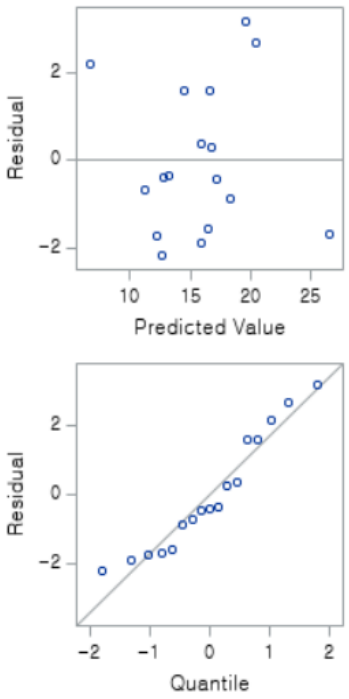
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	307.44350	153.72175	47.43	<.0001
Error	14	45.37415	3.24101		
Corrected Total	16	352.81765			

Root MSE	1.80028	R-Square	0.8714
Dependent Mean	15.71176	Adj R-Sq	0.8530
Coeff Var	11.45817		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	15.77936	0.50700	31.12	<.0001
z	1	0.50186	0.05353	9.37	<.0001
z2	1	-0.00084671	0.00353	-0.24	0.8141



Z2의 p값이 매우 크므로 설명변수 변환의미가 없다고 볼 수 있고 z2와 Y의 잔차산점도 및 정규확률그림은 위 z와 Y의 잔차산점도 및 정규확률그림과 일치한다. 따라서 2차 다항회귀모형은 적절하지않다.

The REG Procedure
Model: MODEL3
Dependent Variable: Y

Number of Observations Read	17
Number of Observations Used	17

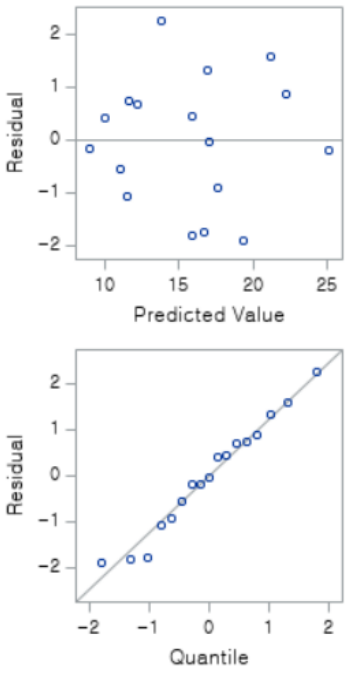
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	329.07938	109.69313	60.07	<.0001
Error	13	23.73827	1.82602		
Corrected Total	16	352.81765			

Root MSE	1.35130	R-Square	0.9327
Dependent Mean	15.71176	Adj R-Sq	0.9172
Coeff Var	8.60058		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	15.62058	0.38334	40.75	<.0001
z	1	0.72894	0.07724	9.44	<.0001
z2	1	0.00510	0.00317	1.61	0.1310
z3	1	-0.00083901	0.00024374	-3.44	0.0044



p값이 모두 유의수준 0.05보다 작으므로 y와 z3는 상당히 유의적이다라는 것을 알 수 있다. 또한 잔차산점도와 정규확률그림은 위의 잔차산점도와 정규확률그림들보다 많이 개선됨을 알 수있다. 따라서 3차 다항회귀모형은 적절하다.

The REG Procedure
Model: MODEL4
Dependent Variable: Y

Number of Observations Read	17
Number of Observations Used	17

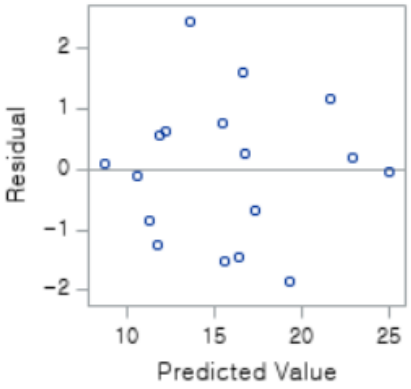
Analysis of Variance

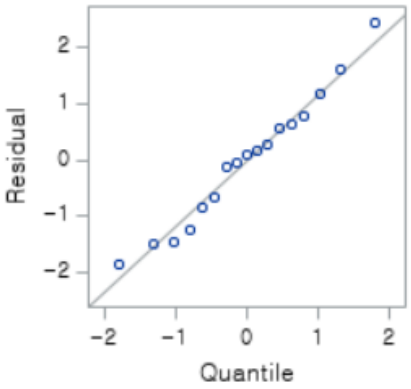
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	330.76246	82.69061	44.99	<.0001
Error	12	22.05519	1.83793		
Corrected Total	16	352.81765			

Root MSE	1.35570	R-Square	0.9375
Dependent Mean	15.71176	Adj R-Sq	0.9167
Coeff Var	8.62859		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	15.31577	0.49937	30.67	<.0001
z	1	0.71790	0.07835	9.16	<.0001
z2	1	0.01877	0.01462	1.28	0.2237
z3	1	-0.00064639	0.00031673	-2.04	0.0639
z4	1	-0.00003420	0.00003574	-0.96	0.3575





Z4의 p값이 유의수준 0.05보다 크므로 설명변수 변환이 의미가 없다. 또한 잔차산점도와 정규확률그림도 z3와 Y의 잔차산점도와 정규확률그림과 일치한다. 따라서 4차 다중회귀모형은 적절하지 않다.

7.4.3

p값이 모두 0.05보다 작고 잔차산점도와 정규확률그림 모두 많이 개선된 3차 다중회귀모형이 가장 적절하다.

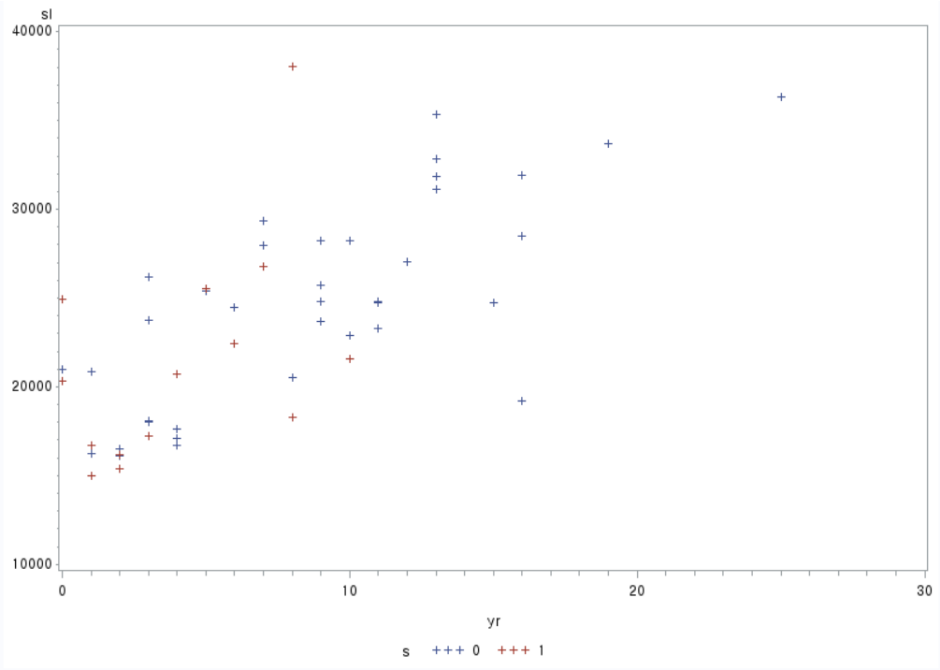
7.4.4

$$y = \beta_0 + \beta_1(x-\bar{x}) + \beta_2(x-\bar{x})^2 + \beta_3(x-\bar{x})^3 + \varepsilon$$
$$= \beta_0 + \beta_1(x-30.1) + \beta_2(x-30.1)^2 + \beta_3(x-30.1)^3 + \varepsilon$$

7.8

7.8.1

성별 sx를 s라는 변수로 두었다.



7.8.2

The REG Procedure
Model: MODEL1
Dependent Variable: sl

Number of Observations Read	52
Number of Observations Used	52

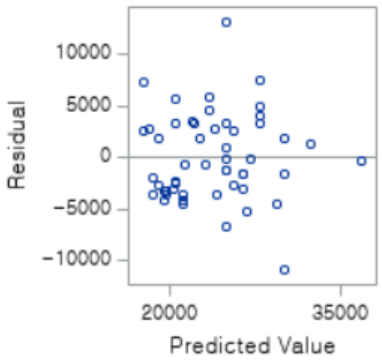
Analysis of Variance

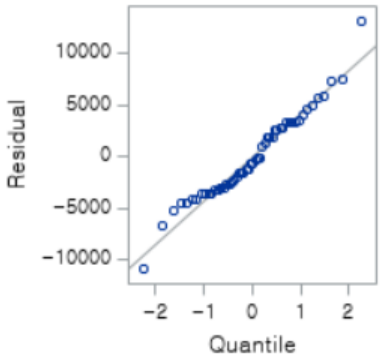
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	880635387	293545129	15.57	<.0001
Error	48	905094471	18856135		
Corrected Total	51	1785729858			

Root MSE	4342.36512	R-Square	0.4932
Dependent Mean	23798	Adj R-Sq	0.4615
Coeff Var	18.24703		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	18223	1308.63164	13.92	<.0001
yr	1	741.02357	126.23109	5.87	<.0001
s	1	-570.75433	2297.23979	-0.25	0.8048
xs	1	169.05347	386.95416	0.44	0.6642





설명변수 yr, 지시변수 s ,설명변수와 지시변수의 교호작용을 포함하는 변수 xs를 만들어서 회귀분석을 진행하였다. p값은 유의수준 0.05보다 크므로 유의적이지않고, 정규확률그림은 별 문제가 없

지만, 잔차산점도는 등

분산성에 약간의 문제가 있어보인다. 따라서 더 정확한 판단을 위해 스코어검정을 실시하려한다. e^2/(SSE/52)=e^2/17405662.9를 u라는 새로운 반응변수로 만들어 설명변수 sl과 yr에 각각 회귀시키면 아래와 같이 u를 yr에 sse/2는 별로 크지않지만 u를 sl로 회귀시켰을 때 sse/2 즉 약 4.2는 자유도가 1인 카이제곱분포에서 매우크므로 오차의 분산은 반응변수에 영향을 받는다는 것을 알 수있다. 따라서 등분산성오차에 문제가 있다는 것을 알 수 있다.

The REG Procedure
Model: MODEL1
Dependent Variable: u

Number of Observations Read	52
Number of Observations Used	52

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8.53909	8.53909	3.15	0.0822
Error	50	135.68268	2.71365		
Corrected Total	51	144.22177			

Root MSE	1.64732	R-Square	0.0592
Dependent Mean	1.00000	Adj R-Sq	0.0404
Coeff Var	164.73171		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.64563	0.95540	-0.68	0.5023
sl	1	0.00006915	0.00003898	1.77	0.0822

The REG Procedure
Model: MODEL2
Dependent Variable: u

Number of Observations Read	52
Number of Observations Used	52

Analysis of Variance

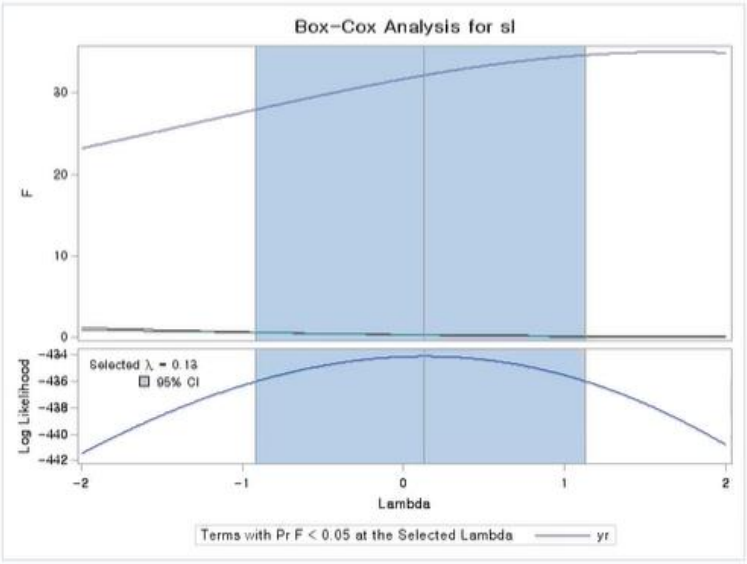
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.80761	0.80761	0.28	0.5980
Error	50	143.41416	2.86828		
Corrected Total	51	144.22177			

Root MSE	1.69360	R-Square	0.0056
Dependent Mean	1.00000	Adj R-Sq	-0.0143
Coeff Var	169.36007		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.82908	0.39865	2.08	0.0427
yr	1	0.02285	0.04306	0.53	0.5980

따라서 boxcox 방법을 사용해서 반응변수의 적절한 변환을 생각해보려한다.



람다에 대한 최대강도 추정값이 0.13으로 0에 가까우므로 로그변환을 취하는 것을 제시하고있으나 신뢰구간이 1을 포함하므로 꼭 반응변수 변환을 할 필요는 없지만 변환을 시도해보면 다음과같다.

SAS 시스템

The REG Procedure
Model: MODEL1
Dependent Variable: lnsl

Number of Observations Read	52
Number of Observations Used	52

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.52042	0.50681	15.13	<.0001
Error	48	1.60819	0.03350		
Corrected Total	51	3.12861			

Root MSE	0.18304	R-Square	0.4860
Dependent Mean	10.04731	Adj R-Sq	0.4538
Coeff Var	1.82179		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.82654	0.05516	178.14	<.0001
yr	1	0.02997	0.00532	5.63	<.0001
s	1	-0.05156	0.09683	-0.53	0.5969
xs	1	0.00957	0.01631	0.59	0.5602

왼쪽결과에서 나온 e^2/(SSE/2)=e^2/0.0309267를 u라는 새로운 반응변수로 두고 다시 sl과 yr에 회귀분석을 해보면 다음과 같다.

SAS 시스템

The REG Procedure
Model: MODEL1
Dependent Variable: lnsl

Number of Observations Read	52
Number of Observations Used	52

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.52042	0.50681	15.13	<.0001
Error	48	1.60819	0.03350		
Corrected Total	51	3.12861			

Root MSE	0.18304	R-Square	0.4860
Dependent Mean	10.04731	Adj R-Sq	0.4538
Coeff Var	1.82179		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.82654	0.05516	178.14	<.0001
yr	1	0.02997	0.00532	5.63	<.0001
s	1	-0.05156	0.09683	-0.53	0.5969
xs	1	0.00957	0.01631	0.59	0.5602

The REG Procedure
Model: MODEL2
Dependent Variable: u

Number of Observations Read	52
Number of Observations Used	52

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.19517	0.19517	0.10	0.7474
Error	50	93.05188	1.86104		
Corrected Total	51	93.24705			

Root MSE	1.36420	R-Square	0.0021
Dependent Mean	1.00000	Adj R-Sq	-0.0179
Coeff Var	136.41968		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.08403	0.32111	3.38	0.0014
yr	1	-0.01123	0.03468	-0.32	0.7474

다시 스코어검정을 실시해보면, yr,sl을 각각 설명변수로 뒀을 때 SSR이 매우 작게 나왔으므로 등분산성 문제는 해결되었다고 볼 수 있다.

따라서 반응변수를 로그로 취하고 설명변수 yr, 지시변수 s ,설명변수와 지시변수의 교호작용을 포함하는 변수 xs로 회귀분석을 진행하였다. 이 결과는 7.8.3에서 설명하겠다.

7.8.3

The REG Procedure
Model: MODEL3
Dependent Variable: lnsl

Number of Observations Read	52
Number of Observations Used	52

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.52042	0.50681	15.13	<.0001
Error	48	1.60819	0.03350		
Corrected Total	51	3.12861			

Root MSE	0.18304	R-Square	0.4860
Dependent Mean	10.04731	Adj R-Sq	0.4538
Coeff Var	1.82179		

Parameter Estimates

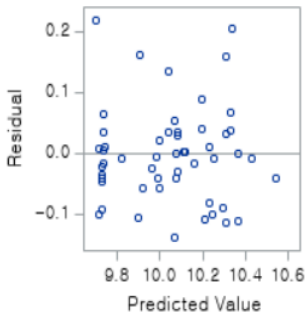
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.82654	0.05516	178.14	<.0001
yr	1	0.02997	0.00532	5.63	<.0001
s	1	-0.05156	0.09683	-0.53	0.5969
xs	1	0.00957	0.01631	0.59	0.5602

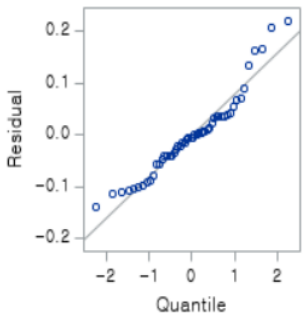
Xs, s(성별)에 대한 p값은 0.05이상으로 유의적이지않다. 그러나 yr(월급액)과 근속연수의 관계는 p값이 유의수준 0.05보다 작으므로 유의적이라고 할 수가 있다. 따라서 성별에 따라 월급액과 근속연수의 관계가 다르다고 볼 수 없다.

7.8.4.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.11059	0.09596	105.36	<.0001
yr	1	0.01687	0.00689	2.45	0.0193
s	1	0.00755	0.11942	0.06	0.9499
r1	1	-0.21200	0.11616	-1.83	0.0761
r2	1	-0.09104	0.09683	-0.94	0.3532
d	1	-0.04777	0.09624	-0.50	0.6226
xs	1	0.01208	0.00985	1.23	0.2277
xr1	1	-0.01569	0.00980	-1.60	0.1179
xr2	1	-0.01096	0.00750	-1.46	0.1525
xd	1	0.00224	0.00687	0.33	0.7465
sr1	1	-0.05384	0.07804	-0.69	0.4946
sr2	1	-0.13884	0.12334	-1.13	0.2675
sd	1	0.01928	0.09407	0.20	0.8388
dr1	1	-0.12559	0.09739	-1.29	0.2052
dr2	1	0.02005	0.08085	0.25	0.8055





이는 월급액과 근속연수의 회귀모형에 성별, 직위 ,최종학력을 모두 포함시킨 경우 회귀모형이다. 잔차산점도는 문제가 없다고 판단되며 정규확률그림은 거의 직선에 가깝지만 약간의 문제가 있다고 볼수있다.

The REG Procedure
Model: MODEL1
Dependent Variable: lnsl

Number of Observations Read	52
Number of Observations Used	52

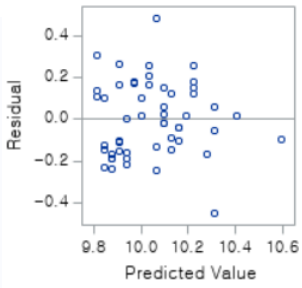
Analysis of Variance

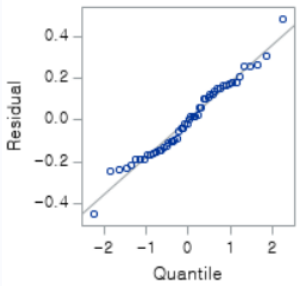
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.50835	1.50835	46.55	<.0001
Error	50	1.62026	0.03241		
Corrected Total	51	3.12861			

Root MSE	0.18001	R-Square	0.4821
Dependent Mean	10.04731	Adj R-Sq	0.4718
Coeff Var	1.79167		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.81372	0.04237	231.61	<.0001
yr	1	0.03123	0.00458	6.82	<.0001





이는 월급액과 근속연수의 단순회귀모형이다. 잔차산점도는 이분산성으로 판단되며, 정규확률그림은 직선에 가깝지만 아직 문제가 남아 있다고 볼 수있다.

따라서 두 모형을 비교했을 때 월급액과 근속연수의 회귀모형에 성별, 직위, 최종학력을 포함시킨 결과가 단순회귀모형보다 잔차산점도와 정규확률그림을 봤을 때 더 적절하다고 할 수 있다.

7.8.5

7.8.4의 월급액과 근속연수의 회귀모형에 성별, 직위, 최종학력을 포함시킨 모형을 봤을 때 t의 절댓값이 가장 작고 p값이 유의수준0.05보다 크다면 가장 큰 변수가 가장 덜 유의적이므로 제거해야하고 모든 변수가 유의적이라면 그게 최적모형이된다. S의 t값이 가장 작고 p가 유의수준 0.05보다 크고 모든 변수중 p값이 가장 크므로 s를 제거 해야한다.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.09837	0.08131	124.20	<.0001
yr	1	0.01735	0.00652	2.66	0.0114
s	1	0.01732	0.11134	0.16	0.8772
r1	1	-0.20401	0.11021	-1.85	0.0719
r2	1	-0.07771	0.07953	-0.98	0.3347
d	1	-0.03235	0.07254	-0.45	0.6582
xs	1	0.01155	0.00949	1.22	0.2313
xr1	1	-0.01581	0.00966	-1.64	0.1102
xr2	1	-0.01110	0.00738	-1.50	0.1409
xd	1	0.00165	0.00637	0.26	0.7971
sr1	1	-0.05520	0.07688	-0.72	0.4771
sr2	1	-0.14873	0.11525	-1.29	0.2047
sd	1	0.01138	0.08742	0.13	0.8971
dr1	1	-0.13580	0.08718	-1.56	0.1276

s를 제거하면 이와 같은 모형이다.

dr2의 t값이 가장 작고 p가 유의수준 0.05보다 크고 모든 변수중 p값이 가장 크므로 dr2를 제거 해야한다.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.10631	0.06249	161.73	<.0001
yr	1	0.01680	0.00541	3.11	0.0035
r1	1	-0.20810	0.10568	-1.97	0.0561
r2	1	-0.08304	0.07085	-1.17	0.2483
d	1	-0.03660	0.06634	-0.55	0.5843
xs	1	0.01231	0.00802	1.54	0.1327
xr1	1	-0.01558	0.00944	-1.65	0.1067
xr2	1	-0.01078	0.00700	-1.54	0.1314
xd	1	0.00195	0.00600	0.32	0.7474
sr1	1	-0.04851	0.06287	-0.77	0.4451
sr2	1	-0.13671	0.08439	-1.62	0.1133
sd	1	0.02081	0.06217	0.33	0.7396
dr1	1	-0.13565	0.08608	-1.58	0.1231

이는 위 모형에서 dr2를 제거 한 모형이다

xd는 t의 절댓값이 가장 작고, p값이 유의수준 0.05보다 크고 모든 변수와 비교했을 때 p값이 가장크므로 제거해야한다.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.09562	0.05250	192.31	<.0001
yr	1	0.01803	0.00381	4.73	<.0001
r1	1	-0.20428	0.10385	-1.97	0.0561
r2	1	-0.08446	0.06992	-1.21	0.2342
d	1	-0.01849	0.03546	-0.52	0.6049
xs	1	0.01256	0.00789	1.59	0.1195
xr1	1	-0.01619	0.00915	-1.77	0.0844
xr2	1	-0.01064	0.00690	-1.54	0.1311
sr1	1	-0.04728	0.06205	-0.76	0.4506
sr2	1	-0.13268	0.08253	-1.61	0.1158
sd	1	0.01755	0.06066	0.29	0.7739
dr1	1	-0.14268	0.08237	-1.73	0.0910

이는 위 모형에서 xd를 제거 한 모형이다

sd는 t의 절댓값이 가장 작고, p값이 유의수준 0.05보다 크고 모든 변수와 비교했을 때 p값이 가장크므로 제거해야한다.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.09852	0.05095	198.20	<.0001
yr	1	0.01782	0.00370	4.82	<.0001
r1	1	-0.20806	0.10186	-2.04	0.0476
r2	1	-0.08787	0.06815	-1.29	0.2045
d	1	-0.01727	0.03481	-0.50	0.6225
xs	1	0.01365	0.00685	1.99	0.0529
xr1	1	-0.01644	0.00900	-1.83	0.0752
xr2	1	-0.01042	0.00679	-1.54	0.1322
sr1	1	-0.03706	0.05044	-0.73	0.4667
sr2	1	-0.13765	0.07982	-1.72	0.0922
dr1	1	-0.13946	0.08070	-1.73	0.0915

이는 위 모형에서 sd를 제거 한 모형이다.

d는 t의 절댓값이 가장 작고, p값이 유의수준 0.05보다 크고 모든 변수와 비교했을 때 p값이 가장크므로 제거해야한다.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.08748	0.04542	222.09	<.0001
yr	1	0.01769	0.00365	4.84	<.0001
r1	1	-0.19795	0.09890	-2.00	0.0518
r2	1	-0.08508	0.06730	-1.26	0.2131
xs	1	0.01312	0.00670	1.96	0.0569
xr1	1	-0.01613	0.00890	-1.81	0.0772
xr2	1	-0.01016	0.00670	-1.51	0.1373
sr1	1	-0.03501	0.04982	-0.70	0.4861
sr2	1	-0.12743	0.07643	-1.67	0.1029
dr1	1	-0.15667	0.07221	-2.17	0.0357

이는 위 모형에서 d를 제거 한 모형이다.

sr1은 t의 절댓값이 가장 작고, p값이 유의수준 0.05보다 크고 모든 변수와 비교했을 때 p값이 가장크므로 제거해야한다.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.09311	0.04444	227.10	<.0001
yr	1	0.01739	0.00361	4.82	<.0001
r1	1	-0.21600	0.09494	-2.28	0.0279
r2	1	-0.09087	0.06640	-1.37	0.1783
xs	1	0.01068	0.00570	1.87	0.0678
xr1	1	-0.01480	0.00865	-1.71	0.0942
xr2	1	-0.00984	0.00665	-1.48	0.1463
sr2	1	-0.11520	0.07398	-1.56	0.1268
dr1	1	-0.16151	0.07145	-2.26	0.0289

이는 위 모형에서 sr1을 제거 한 모형이다. r2는 t의 절댓값이 가장 작고, p값이 유의수준 0.05보다 크고 모든 변수와 비교했을 때 p값이 가장크므로 제거해야한다.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.05231	0.03328	302.03	<.0001
yr	1	0.02033	0.00292	6.95	<.0001
r1	1	-0.17440	0.09083	-1.92	0.0614
xs	1	0.01227	0.00564	2.18	0.0348
xr1	1	-0.01825	0.00835	-2.18	0.0343
xr2	1	-0.01766	0.00343	-5.15	<.0001
sr2	1	-0.14883	0.07046	-2.11	0.0404
dr1	1	-0.16291	0.07215	-2.26	0.0290

이는 위 모형에서 r2를 제거 한 모형이다. r1는 t의 절댓값이 가장 작고, p값이 유의수준 0.05보다 크고 모든 변수와 비교했을 때 p값이 가장크므로 제거해야한다.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.02845	0.03178	315.51	<.0001
yr	1	0.02204	0.00287	7.69	<.0001
xs	1	0.01348	0.00577	2.34	0.0238
xr1	1	-0.03180	0.00460	-6.91	<.0001
xr2	1	-0.01704	0.00352	-4.85	<.0001
sr2	1	-0.14263	0.07246	-1.97	0.0552
dr1	1	-0.28362	0.03644	-7.78	<.0001

이는 위 모형에서 r1를 제거 한 모형이다. sr2는 t의 절댓값이 가장 작고, p값이 유의수준 0.05보다 크고 모든 변수와 비교했을 때 p값이 가장크므로 제거해야한다.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.01917	0.03240	309.23	<.0001
yr	1	0.02292	0.00292	7.85	<.0001
xs	1	0.00947	0.00556	1.70	0.0951
xr1	1	-0.03085	0.00471	-6.54	<.0001
xr2	1	-0.01821	0.00357	-5.10	<.0001
dr1	1	-0.27429	0.03724	-7.37	<.0001

이는 위 모형에서 sr2를 제거 한 모형이다. xs는 t의 절댓값이 가장 작고, p값이 유의수준 0.05보다 크고 모든 변수와 비교했을 때 p값이 가장크므로 제거해야한다.

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	10.03385	0.03186	314.93	<.0001
yr	1	0.02229	0.00295	7.54	<.0001
xr1	1	-0.02871	0.00464	-6.19	<.0001
xr2	1	-0.01858	0.00364	-5.11	<.0001
dr1	1	-0.28145	0.03775	-7.46	<.0001

이제는 모든 변수의 p값이 유의수준 0.05보다 작으므로 이 모형이 월급액을 반응변수로하는 최적모형이다. 또한 s에 관한 변수는 이 모형의 변수에 아예 포함되어있지않으므로 성별에 의한 월급액에 차이가 없다는 것을 알 수 있다.