Instructor: Don Negri
Office: Smullin 307
Email: dnegri@willamette.edu

Homework #10 & Practice Final

I. FBI Crime Report Analysis

Variable Names:

1. R: Crime rate: # of offenses reported to police per million population
2. Age: The number of males of age 14-24 per 1000 population
3. S: Indicator variable for Southern states (0 = No, 1 = Yes)
4. Ed: Mean # of years of schooling x 10 for persons of age 25 or older
5. Ex0: 1960 per capita expenditure on police by state and local government
6. Ex1: 1959 per capita expenditure on police by state and local government
7. LF: Labor force participation rate per 1000 civilian urban males age 14-24
8. M: The number of males per 1000 females
9. N: State population size in hundred thousands
10. NW: The number of non-whites per 1000 population
11. U1: Unemployment rate of urban males per 1000 of age 14-24
12. U2: Unemployment rate of urban males per 1000 of age 35-39
13. W: Median value of transferable goods and assets or family income in tens of $
14. X: The number of families per 1000 earning below 1/2 the median income

These data are crime-related and demographic statistics for 47 US states in 1960. The data were collected from the FBI's *Uniform Crime Report* and other government agencies to determine how the variable crime rate depends on the other variables measured in the study. **Table shows only 15 of the 47 observations.**

R	Age	S	Ed	Ex0	Ex1	LF	M	N	NW	U1	U2	W	X
79.1	151	1	91	58	56	510	950	33	301	108	41	394	261
163.5	143	0	113	103	95	583	1012	13	102	96	36	557	194
57.8	142	1	89	45	44	533	969	18	219	94	33	318	250
196.9	136	0	121	149	141	577	994	157	80	102	39	673	167
123.4	141	0	121	109	101	591	985	18	30	91	20	578	174
68.2	121	0	110	118	115	547	964	25	44	84	29	689	126
96.3	127	1	111	82	79	519	982	4	139	97	38	620	168
155.5	131	1	109	115	109	542	969	50	179	79	35	472	206
85.6	157	1	90	65	62	553	955	39	286	81	28	421	239
70.5	140	0	118	71	68	632	1029	7	15	100	24	526	174
167.4	124	0	105	121	116	580	966	101	106	77	35	657	170
84.9	134	0	108	75	71	595	972	47	59	83	31	580	172
51.1	128	0	113	67	60	624	972	28	10	77	25	507	206
66.4	135	0	117	62	61	595	986	22	46	77	27	529	190
79.8	152	1	87	57	53	530	986	30	72	92	43	405	264

ANSWER THE FOLLOWING QUESTIONS (from the Regression Table):

1. Before you read the regression table, what are your expectations about the relationship between crime and the independent variables?
2. What is the sample regression line?
3. What variables are statistically significant predictors of crime rates? What variables are not significant? Does the outcome surprise you? Why or why not?
4. Which relationships are statistically significant at the 5% level of significance? At the 10% level of significance? Carefully explain what it means to be statistically significant and statistically insignificant.
5. What is the value of the coefficient of determination (r^2)? Explain its meaning. How would you compute it using the Analysis of Variance Table?
6. Test the hypothesis that the true regression coefficients are all jointly zero.
7. What is the value of the standard error of the regression (S)? Explain how you would compute it from the Analysis of Variance Table. What does this value mean?
8. Regressions are sometimes used to forecast values of the dependent variable corresponding to values of the independent variable lying outside the range of the sample data. This is known as extrapolation. Use the results to demonstrate the problems that are encountered when extrapolating.
9. What do we learn about the determinants of crime? What does the regression suggest about public policy towards crime?

Crime
Regression

Regression Statistics	
Multiple R	0.877061028
R Square	0.769236046
Adjusted R Square	0.678329034
Standard Error	21.9356502
Observations	47

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	52930.57586	4071.582758	8.461789993	3.68552E-07
Residual	33	15878.70074	481.1727497		
Total	46	68809.2766			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-691.8375879	155.8879181	-4.438044952	9.55789E-05	-1008.99394	-374.6812355
Age	1.039809653	0.422708293	2.459875216	0.019306419	0.179803169	1.899816136
S	-8.308312889	14.91158757	-0.557171586	0.58116995	-38.64616575	22.02953997
Ed	1.80160106	0.649650398	2.773185493	0.009059759	0.479877393	3.123324727
Ex0	1.607818377	1.058666975	1.518719687	0.138356548	-0.546055768	3.761692522
Ex1	-0.667258285	1.148773449	-0.580844104	0.565291613	-3.004455429	1.669938859
LF	-0.041031047	0.15347688	-0.267343508	0.790868482	-0.353282105	0.27122001
M	0.164794968	0.209931929	0.784992397	0.438057576	-0.26231475	0.591904686
N	-0.041276887	0.129516066	-0.318700899	0.751961799	-0.304779302	0.222225529
NW	0.007174688	0.063867197	0.112337606	0.911236296	-0.122764101	0.137113477
U1	-0.601675298	0.43715449	-1.376344776	0.177983068	-1.491072791	0.287722194
U2	1.792262901	0.856111433	2.093492541	0.044069376	0.050491102	3.534034699
W	0.137358473	0.105830276	1.297912825	0.20331629	-0.077954842	0.352671788
X	0.792932786	0.235085239	3.372958635	0.001912553	0.314648274	1.271217298

II Wage Regression Problem

.

June 2007 Current Population Survey Variable Descriptions

Column	Count	Name	Description
A	6280	WAGE	Hourly Wage – in dollars
B	6280	EXPER	Work Experience in Years
C	6280	EDUC	Years of Education
D	6280	FEMALE	Female Dummy Variable (1=Female, 0 Otherwise)
E	6280	MARRIED	Dummy (1=Married, 0 Otherwise)
F	6280	CHILDREN	Number of Children
G	6280	UNION	Union Dummy Variable (
H	6280	METRO	Dummy (1=Live Major Metropolitan area, 0 Otherwise)
I	6280	BLACK	Dummy (1=Black, 0 Otherwise)
J	6280	HISP	Hispanic Dummy Variable (1=Hispanic, 0 Otherwise)
K	6280	NATIVEAM	Dummy (1=Native American, 0 Otherwise)
L	6280	ASIAN	Dummy Variable (1=Asian, 0 Otherwise)
M	6280	MIXRACE	Dummy (1=Mixed Race, 0 otherwise)
N	6280	PROF	Dummy (1=Professional/Managerial, 0 otherwise)
O	6280	SERVICE	Dummy (1=Service job, 0 otherwise)
P	6280	SALES	Dummy (1=Sales job, 0 otherwise)
Q	6280	AGRIC	Dummy (1=Agricultural, 0 otherwise)
R	6280	CONSTRUCT	Dummy (1=Construction job, 0 Otherwise)
S	6280	XPORT	Dummy (1=Transportation job, 0 otherwise)
T	6280	MALE	Dummy (1=Male, 0 otherwise)
U	6280	AGE	Age of Respondent in Years
V	6280	HRSWRK	Hours worked per week
W	6280	BACHELORS	Dummy (1=Bachelor's highest, 0 otherwise)
X	6280	MASTERS	Dummy (1=Master's highest, 0 Otherwise)
Y	6280	PROFDEG	Dummy (1=Professional Degree highest, 0 Otherwise)
Z	6280	PHD	Dummy (1=Highest degree Ph.D., 0 Otherwise)
AA	6280	WHITE	Dummy (1=White, 0 otherwise)

EXCEL OUTPUT

Wage Regression Excel Output

Regression Statistics	
Multiple R	0.535616771
R Square	0.286885325
Adjusted R Square	0.285063541
Standard Error	6.3323956
Observations	6280

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	16	101033.9771	6314.623567	157.4749174	0
Residual	6263	251141.5027	40.09923403		
Total	6279	352175.4798			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1.099338837	0.522948241	2.10219435	0.035576074	0.07418103	2.124496645	0.07418103	2.124496645
exper	0.090102798	0.006167292	14.60978275	1.46451E-47	0.078012791	0.102192804	0.078012791	0.102192804
educ	0.826064177	0.037588685	21.97640539	3.5343E-103	0.75237747	0.899750884	0.75237747	0.899750884
female	-1.720588571	0.184575517	-9.321867827	1.55155E-20	-2.082419852	-1.35875729	-2.082419852	-1.35875729
married	1.018721942	0.194392623	5.240538064	1.65344E-07	0.637645766	1.399798117	0.637645766	1.399798117
children	0.383232001	0.090476656	4.235700325	2.31105E-05	0.205866742	0.56059726	0.205866742	0.56059726
union	0.71983954	0.773780913	0.930288571	0.352257549	-0.797036285	2.236715365	-0.797036285	2.236715365
metro	0.984239463	0.189450169	5.195241928	2.10915E-07	0.612852191	1.355626734	0.612852191	1.355626734
black	-1.073103616	0.281709218	-3.809259858	0.000140717	-1.625350247	-0.520856985	-1.625350247	-0.520856985
hisp	-0.049412025	0.245779767	-0.201041873	0.840672361	-0.531224616	0.432400566	-0.531224616	0.432400566
NativeAm	-1.423415898	0.838351092	-1.697875641	0.089580913	-3.066871407	0.22003961	-3.066871407	0.22003961
asian	0.027878658	0.455787633	0.061165895	0.951229043	-0.865621335	0.921378652	-0.865621335	0.921378652
mixrace	0.120938828	0.560983909	0.215583416	0.829319504	-0.978781927	1.220659582	-0.978781927	1.220659582
prof	3.622784702	0.2767227	13.09175107	1.17684E-38	3.080313355	4.165256048	3.080313355	4.165256048
service	-2.881514825	0.243417226	-11.83776049	5.45056E-32	-3.358696027	-2.404333623	-3.358696027	-2.404333623
sales	-1.817567306	0.240181082	-7.567487377	4.35349E-14	-2.288404556	-1.346730057	-2.288404556	-1.346730057
agric	-2.777569734	0.760014776	-3.654625966	0.000259665	-4.267459211	-1.287680257	-4.267459211	-1.287680257

Omitted: White & transportatioin

ANSWER THE FOLLOWING QUESTIONS:

10. What is the sample regression line?
11. Are education and experience statistically significant predictors of wages? What is the relationship between wage and education? experience? (That is, what is the effect on wage of an additional year of education? Experience?)
12. What is the impact on wages of Gender, Race, Union Membership, working in (a) sales, (b) professional/management and (d) agriculture? Are these relationships statistically significant? At what level of significance?
13. Which relationships are statistically significant at the 5% level of significance? At the 10% level of significance? Carefully explain what it means to be statistically significant and statistically insignificant.
14. What is the value of the coefficient of determination (r^2)? Explain its meaning. How would you compute it using the Analysis of Variance Table?
15. Test the hypothesis that the true regression coefficients are all jointly zero.
16. What is the value of the standard error of the regression (S_e)? Explain how you would compute it from the Analysis of Variance Table. What does this value mean?
17. Estimate the wage for an unmarried, white women who is a professional, non-union, 5 years of experience, no children, lives in a metropolitan area and a 16 years of education. Calculate a 90% and a 95% confidence interval for the conditional mean (i.e., population regression line).
18. Regressions are sometimes used to forecast values of the dependent variable corresponding to values of the independent variable lying outside the range of the sample data. This is known as extrapolation. Use the results to demonstrate the problems that are encountered when extrapolating.
19. Earlier in the semester we compared wages of men and women. We rejected the null hypothesis that wages for men and women were the same at the .01 level of significance using the test for the difference in two population means. We now have used regression analysis to study wages in the United States. What new conclusions can we draw from this analysis? What is the story told by the new analysis? Carefully explain.
20. What do we learn from the regression about the impact of job classification? Marriage? Race? Carefully explain

III. The Value of Econometrics

Data: Econometrics experience and salary
students in the economics department at large state school

Variables

salary	starting annual salary after graduation
GPA	Grade point average for student
metrics	dummy for having taken econometrics course
Female	dummy for female

Partial list of 50 Observations

Salary	*GPA*	*Metrics*	*Female*
29555	3.77	0	1
27958	3.4	0	0
27230	2.74	0	0
31070	3.88	0	1
27577	2.7	0	1
30007	3.36	0	0
28988	2.57	0	0
32655	3.81	0	0
29310	2.7	0	0
29145	3.29	0	1
26142	2.97	0	0
29460	2.16	0	0
28390	2.84	0	1
28574	2.53	0	0
28760	2.79	0	0
29665	3.81	0	0
27481	2.18	0	0
29186	2.25	0	0
29302	3.44	0	1
28917	3.04	0	1
32839	3.78	0	0
27277	2.03	0	0
26531	2.15	0	1
32537	3.89	0	0
28594	2.66	0	0
28268	3.07	0	1
29249	2.34	0	0
28525	3.89	0	0

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.859205845
R Square	0.738234683
Adjusted R Square	0.721163032
Standard Error	1446.330273
Observations	50

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	271378305.8	90459435.25	43.243309	1.94782E-13
Residual	46	96226077.92	2091871.259		
Total	49	367604383.7			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*
Intercept	24241.7208	1090.681444	22.22621549	3.0548E-26	22046.29295	26437.14864	22046.29295
GPA	1657.760819	356.4192782	4.651153628	2.80502E-05	940.3260337	2375.195604	940.3260337
Metrics	5023.513894	460.4685696	10.90956957	2.37626E-14	4096.638752	5950.389037	4096.638752
Female	-204.795086	419.6012905	-0.48807068	0.62781824	-1049.40866	639.8184915	-1049.40866

21. What is the sample regression line?
22. Are GPA, Female and Econometrics statistically significant predictors of salary? What is the relationship between salary and GPA? Female? Metrics? (That is, what is the effect on Salary of an additional point GPA or taking Econometrics or being Female?)
23. What is the value of the coefficient of determination (r^2)?
24. Test the hypothesis that the true regression coefficients are all jointly zero.
25. What is the value of the standard error of the regression (S)? Explain how you would compute it from the Analysis of Variance Table. What does this value mean?
26. Estimate the salary for a male with a 3.5 GPA and no econometrics. A female with a 3.5 GPA and with Econometrics.
27. Show how to calculate a 95% confidence interval for the coefficient on GPA. Interpret its meaning.
28. What do we learn from the regression about the value of taking econometrics?
29. Do you see any problems with this regression?

Additional food for Thought:

30. What is OLS? How is it used to compute the sample regression line?
31. There are three properties of the estimated coefficients that we have discussed in class, Unbiasedness, efficiency and consistency. Explain each in the simplest terms possible.