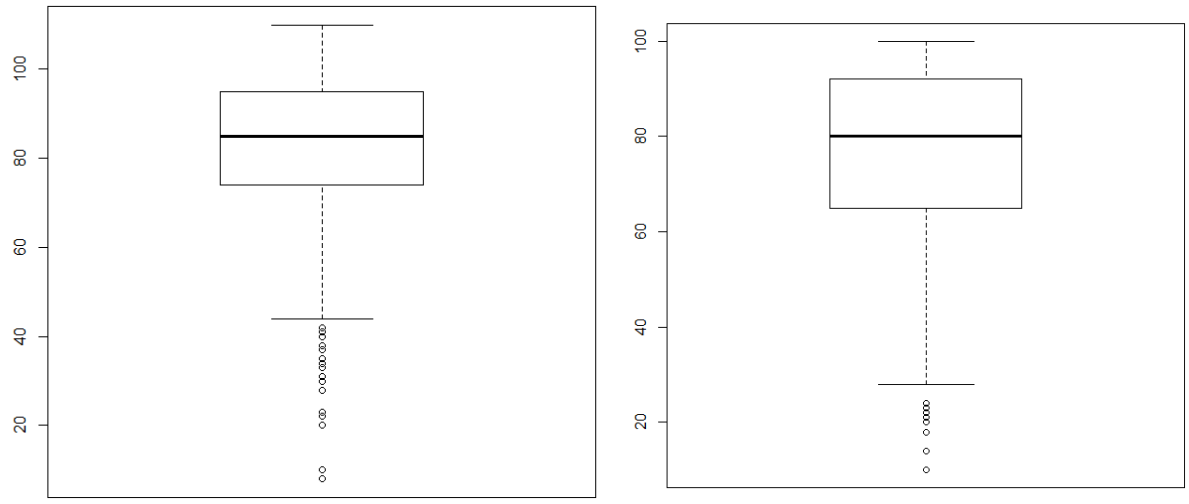
Spring 2016 Statistical Data Analysis with R

Prepared by Ramjee Sharma

Overall Descriptive Statistics

	StudentsGrade	InstructorsGrade	diff
median	85.0000000	80.0000000	-2.0000000
mean	82.4940797	76.6420883	-5.8519914
SE.mean	0.5442731	0.6508337	0.5992326
CI.mean.0.95	1.0681488	1.2772764	1.1760081
var	275.2006653	393.5099616	333.5850203
std.dev	16.5891731	19.8370855	18.2643100
coef.var	0.2010953	0.2588276	-3.1210418

Box Plot of students grades and Instructor's grades



Inferential Statistics

Welch Two Sample t-test

t = -6.8975, df = 1799.7, p-value = 7.302e-12

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: (-7.515985, -4.187998)

Categorical Analysis

Only A students

	StudentsGrade	InstructorsGrade	diff
median	95.0000000	97.00000000	0.0000000
mean	91.0229508	96.24754098	5.2245902
SE.mean	0.6267584	0.22526613	0.5949653
CI.mean.0.95	1.2333341	0.44327828	1.1707715
var	119.8119715	15.47717321	107.9650183
std.dev	10.9458655	3.93410386	10.3906217
coef.var	0.1202539	0.04087485	1.9887917

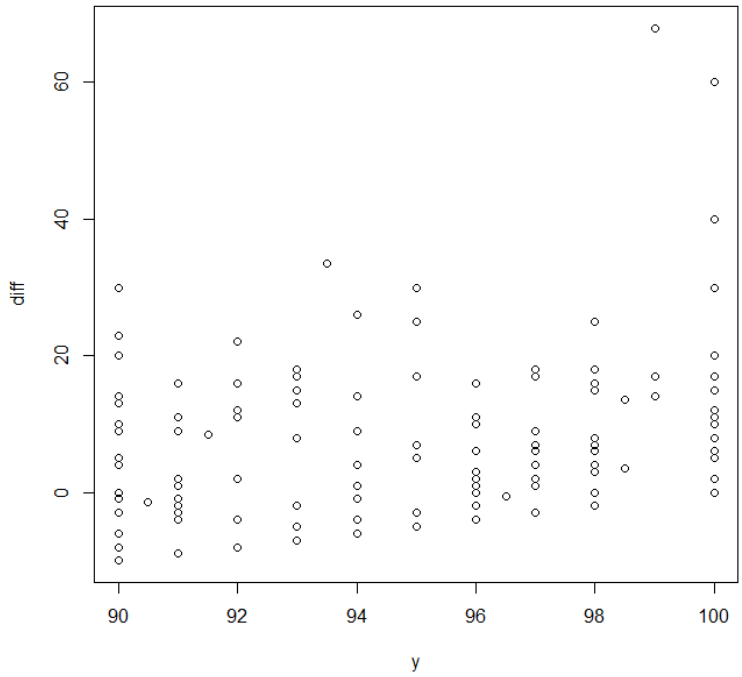
Welch Two Sample t-test

t = 7.8446, df = 381.25, p-value = 4.415e-14

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: 3.915075 6.534105

Scatter plot (Instructor's grades versus differences)



Only B students

StudentsGrade	InstructorsGrade	diff	
median	86.0000000	83.00000000	-1.0000000
mean	84.7203791	83.30331754	-1.4170616
SE.mean	0.8946858	0.22130324	0.8773305
CI.mean.0.95	1.7637163	0.43626056	1.7295034
var	168.8976303	10.33375085	162.4085647
std.dev	12.9960621	3.21461519	12.7439619
coef.var	0.1533995	0.03858928	-8.9932306

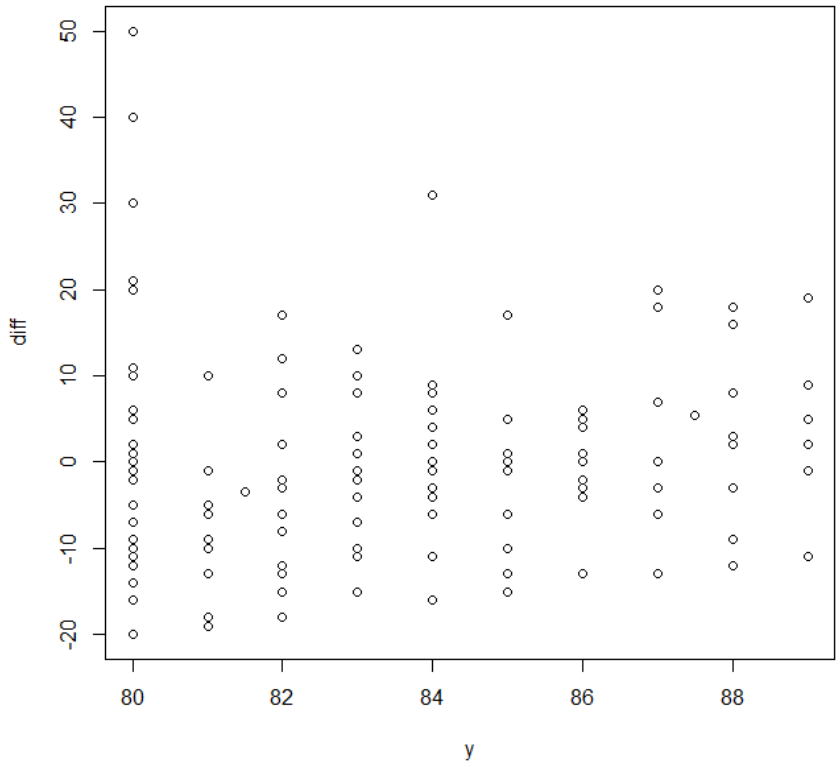
Welch Two Sample t-test

t = -1.5375, df = 235.6, p-value = 0.1255

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: (-3.2327886, 0.3986654)

Scatter plot (instructor's grade versus differences)



A and B students

	StudentsGrade	InstructorsGrade	diff
median	90.0000000	90.25000000	0.0000000
mean	88.4457364	90.95445736	2.5087209
SE.mean	0.5377503	0.32327827	0.5220678
CI.mean.0.95	1.0564541	0.63510635	1.0256445
var	149.2145255	53.92656262	140.6382733
std.dev	12.2153398	7.34347075	11.8591009
coef.var	0.1381111	0.08073789	4.7271503

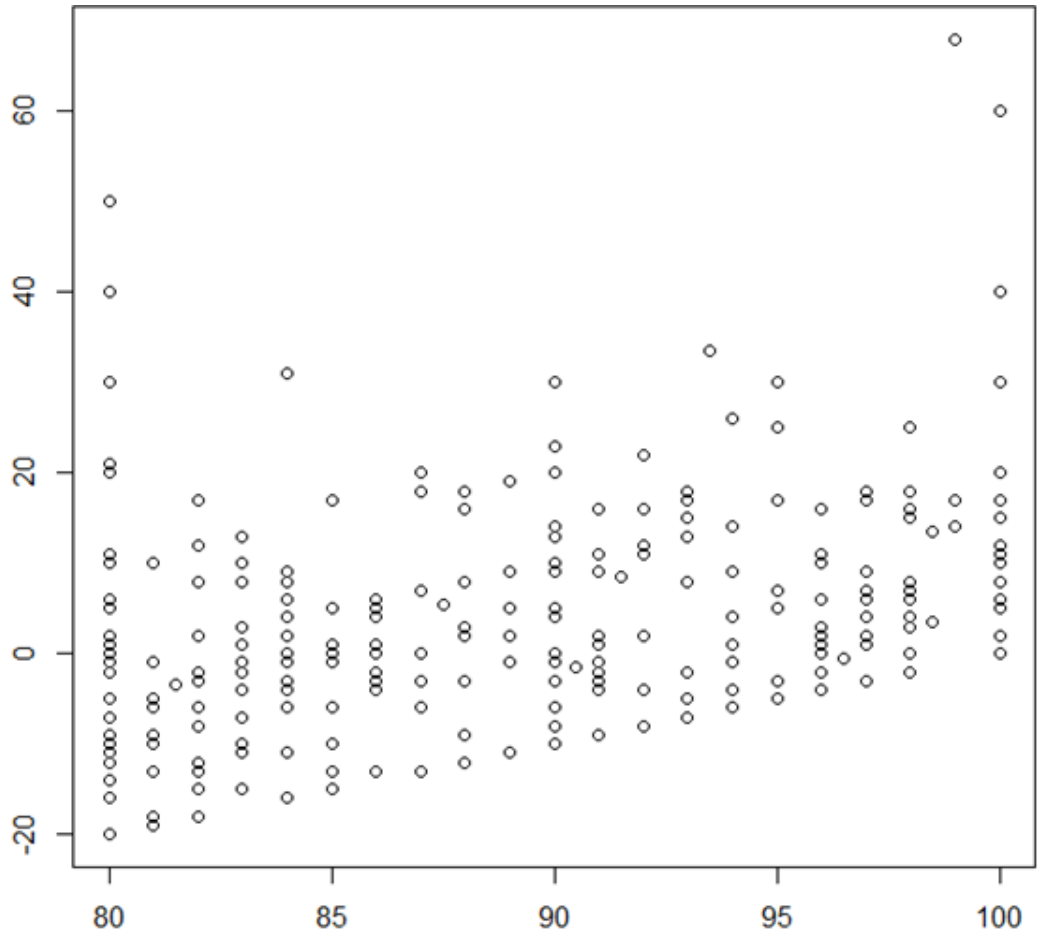
Welch Two Sample t-test

t = 3.9983, df = 844.24, p-value = 6.938e-05

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: (1.277190, 3.740251)

Scatter plot (Instructor's grades versus differences)



Only C

	StudentsGrade	InstructorsGrade	diff
median	84.0000000	74.00000000	-10.000000
mean	82.1811594	73.75362319	-8.427536
SE.mean	1.2516037	0.27517852	1.234440
CI.mean.0.95	2.4749602	0.54414660	2.441021
var	216.1786205	10.44980429	210.290331
std.dev	14.7030140	3.23261570	14.501391
coef.var	0.1789098	0.04382992	-1.720715

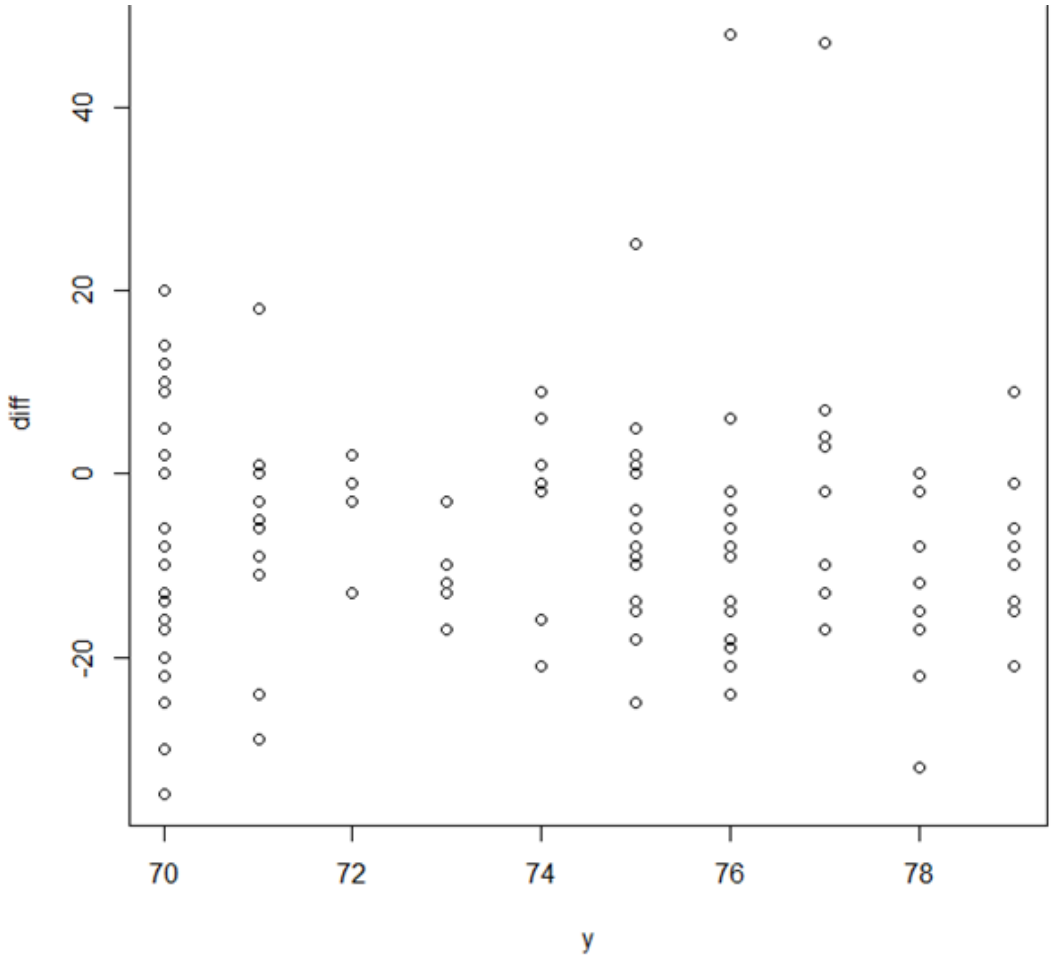
Welch Two Sample t-test

t = -6.5763, df = 150.21, p-value = 7.51e-10

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: (-10.959624, -5.895448)

Scatter plot (Instructor's grades versus differences)



Below C

	StudentsGrade	InstructorsGrade	diff
median	78.0000000	62.0000000	-15.0000000
mean	75.0581114	58.7602906	-16.2978208
SE.mean	0.8986939	0.7652524	0.9575131
CI.mean.0.95	1.7665972	1.5042862	1.8822205
var	333.5597217	241.8574485	378.6513787
std.dev	18.2636174	15.5517667	19.4589665
coef.var	0.2433264	0.2646646	-1.1939613

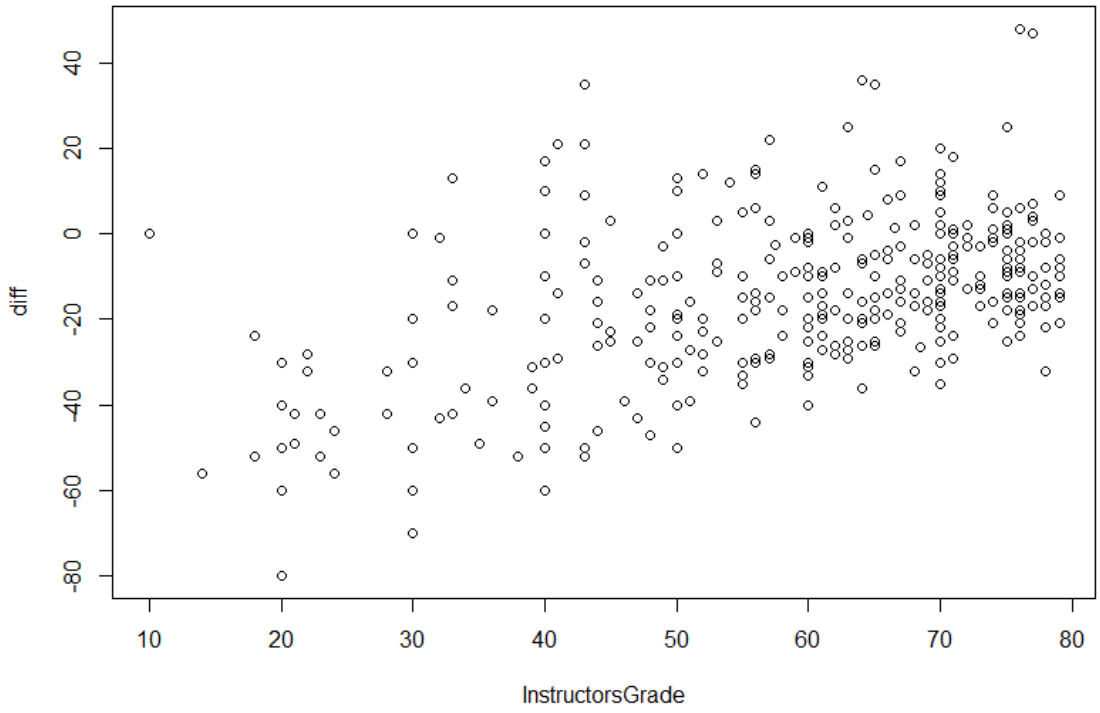
Welch Two Sample t-test

data: y and x

t = -13.807, df = 803.59, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: (-18.61478, -13.98086)

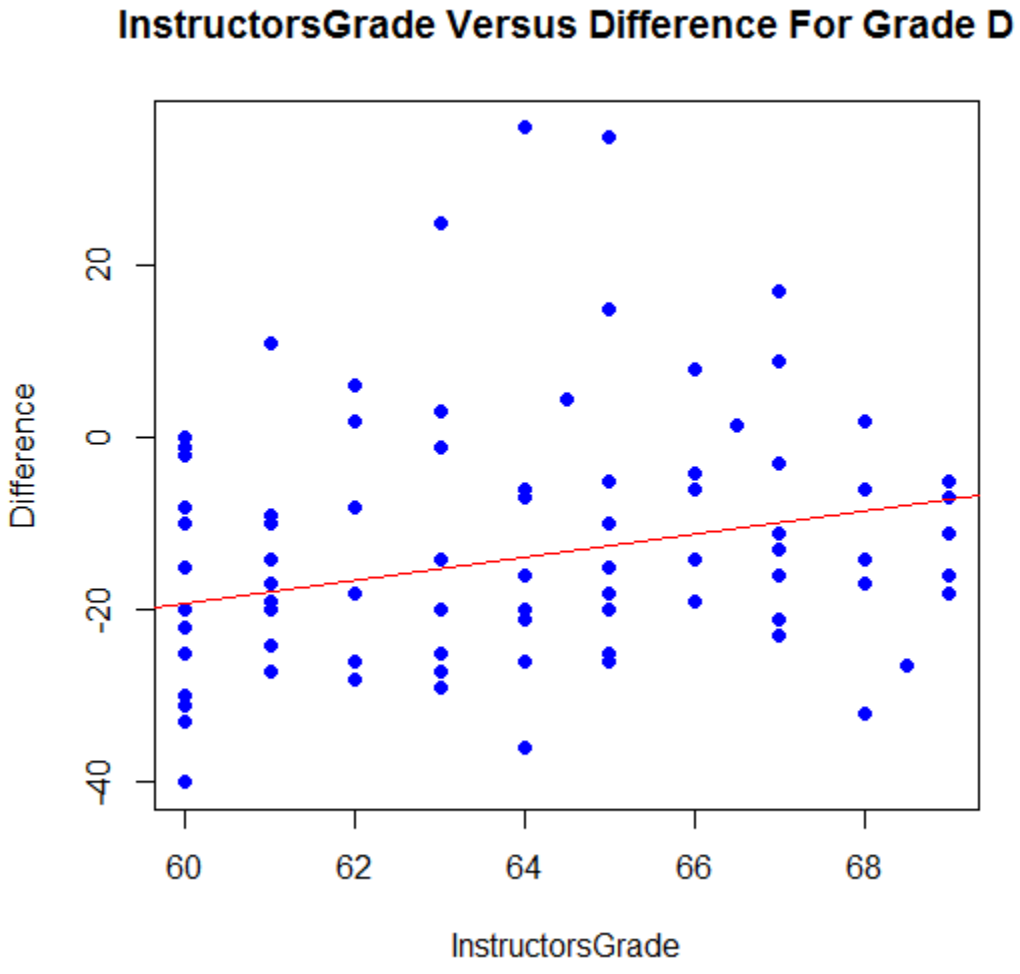


Only D

	StudentsGrade	InstructorsGrade	Difference
nbr.val	114.00	114.000	114.0
nbr.null	0.00	0.000	4.0
nbr.na	0.00	0.000	0.0
min	28.00	60.000	-40.0
max	100.00	69.000	36.0
range	72.00	9.000	76.0
sum	8893.00	7223.500	-1669.5
median	80.00	63.000	-16.0
mean	78.01	63.364	-14.6
SE.mean	1.41	0.279	1.5
CI.mean.0.95	2.80	0.553	2.9
var	227.00	8.891	241.7
std.dev	15.07	2.982	15.5
coef.var	0.19	0.047	-1.1

```
        Paired t-test

data:  StudentsGrade and InstructorsGrade
t = 10, df = 100, p-value <0.0000000000000002
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
  12 Inf
sample estimates:
mean of the differences
                     15
```




```
Residuals:
```

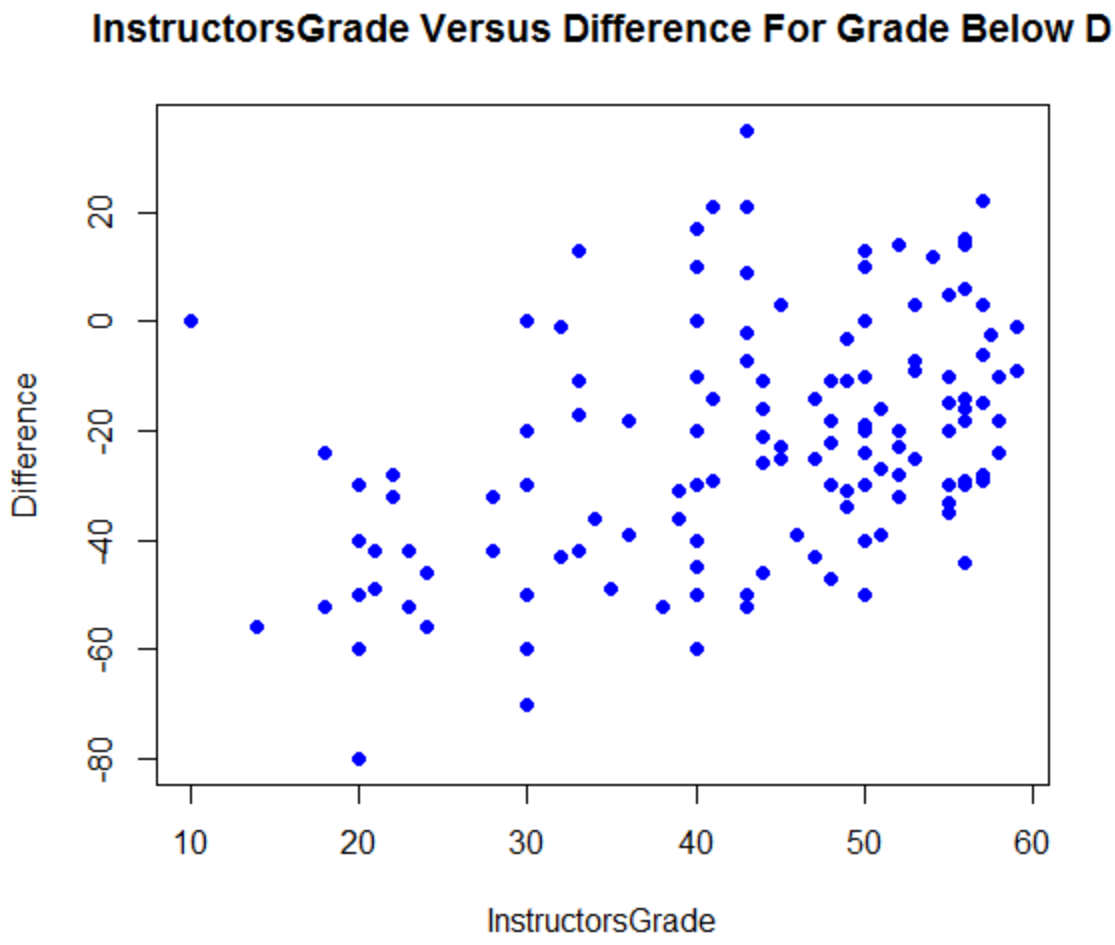
```
   Min     1Q Median     3Q    Max
-23.51 -10.89  -2.21   8.46  49.80

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       -98.782     30.223   -3.27   0.0014 **
InstructorsGrade    1.328      0.476    2.79   0.0063 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15 on 112 degrees of freedom
Multiple R-squared:  0.0649,    Adjusted R-squared:  0.0565
F-statistic: 7.77 on 1 and 112 DF,  p-value: 0.00625
```

Below D

	StudentsGrade	InstructorsGrade	Difference
nbr.val	161.0	161.00	161.00
nbr.null	0.0	0.00	7.00
nbr.na	0.0	0.00	0.00
min	8.0	10.00	-80.00
max	100.0	59.00	35.00
range	92.0	49.00	115.00
sum	10765.0	6866.50	-3898.50
median	70.0	44.00	-25.00
mean	66.9	42.65	-24.21
SE.mean	1.6	0.92	1.77
CI.mean.0.95	3.1	1.82	3.50
var	396.0	137.37	505.82
std.dev	19.9	11.72	22.49
coef.var	0.3	0.27	-0.93




```
        Paired t-test

data:  StudentsGrade and InstructorsGrade
t = 10, df = 200, p-value <0.0000000000000002
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
  21 Inf
sample estimates:
mean of the differences
                     24

Residuals:
   Min     1Q Median     3Q    Max
```

```
 -35.4  -13.6   -2.6   14.6   58.9

Coefficients:
                 Estimate Std. Error t value             Pr(>|t|)
(Intercept)       -62.593      5.944   -10.5 < 0.0000000000000002 ***
InstructorsGrade    0.900      0.134     6.7        0.00000000035 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 20 on 159 degrees of freedom
Multiple R-squared:  0.22,      Adjusted R-squared:  0.215
F-statistic: 44.8 on 1 and 159 DF,  p-value: 0.000000000351
```

D and Below D

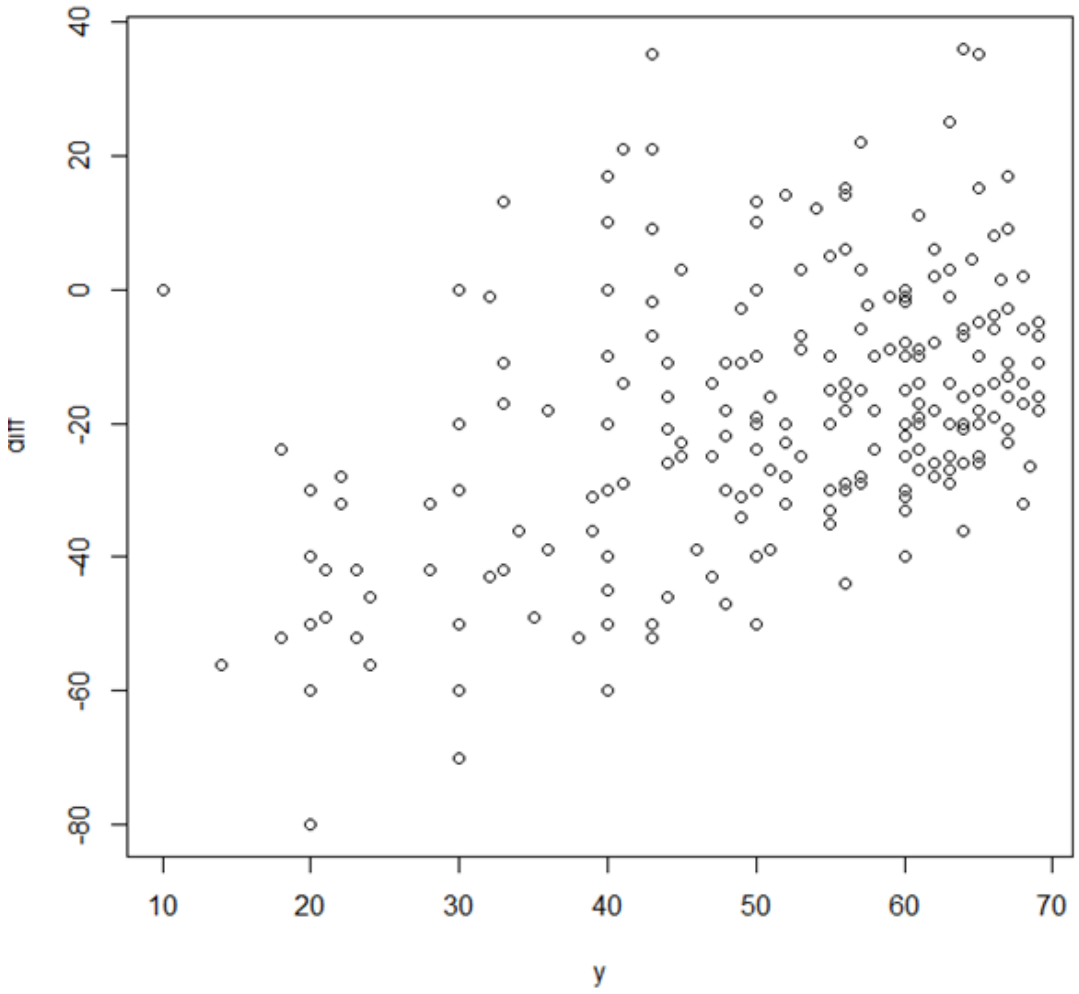
	StudentsGrade	InstructorsGrade	diff
median	72.0000000	55.0000000	-20.000000
mean	71.4836364	51.2363636	-20.247273
SE.mean	1.1363259	0.8277185	1.231942
CI.mean.0.95	2.2370389	1.6294960	2.425274
var	355.0900597	188.4074320	417.361991
std.dev	18.8438335	13.7261587	20.429439
coef.var	0.2636104	0.2678988	-1.008997

Welch Two Sample t-test

t = -14.402, df = 500.89, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: (-23.00932, -17.48522)



Analysis By Courses

Math 1101 Introduction to Mathematical Modeling

	StudentsGrade	InstructorsGrade	diff
median	90.0000000	79.0000000	-6.000000
mean	82.9375000	77.0156250	-5.921875
SE.mean	2.5506710	2.0842251	1.764929
CI.mean.0.95	5.2021277	4.2508051	3.599596
var	208.1895161	139.0078125	99.679183
std.dev	14.4287739	11.7901574	9.983946
coef.var	0.1739717	0.1530879	-1.685943

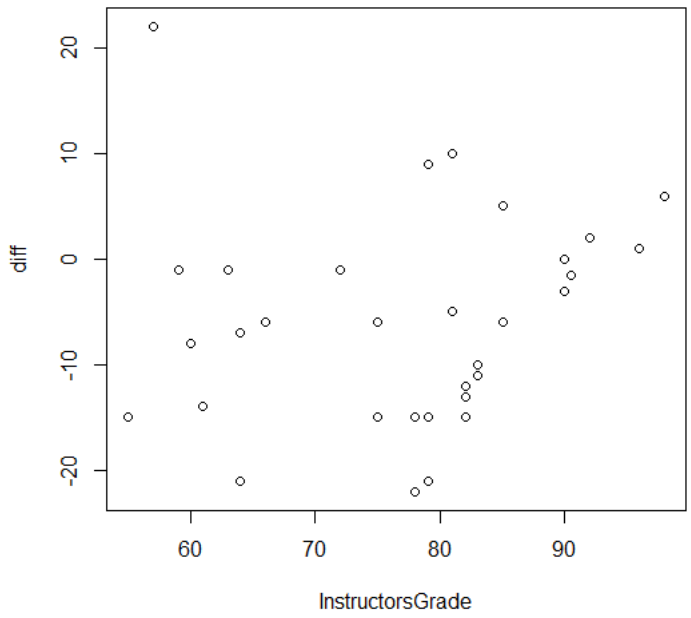
Welch Two Sample t-test

data: InstructorsGrade and StudentsGrade

t = -1.7978, df = 59.632, p-value = 0.07727

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: -12.5115338   0.6677838



Math 1111 College Algebra

StudentsGrade	InstructorsGrade	diff	
median	80.0000000	80.0000000	2.000000
mean	76.5222930	79.0254777	2.503185
SE.mean	1.5301848	1.3837426	1.282829
CI.mean.0.95	3.0225549	2.7332894	2.533956
var	367.6100768	300.6147313	258.366977
std.dev	19.1731603	17.3382448	16.073798
coef.var	0.2505565	0.2194007	6.421339

Welch Two Sample t-test

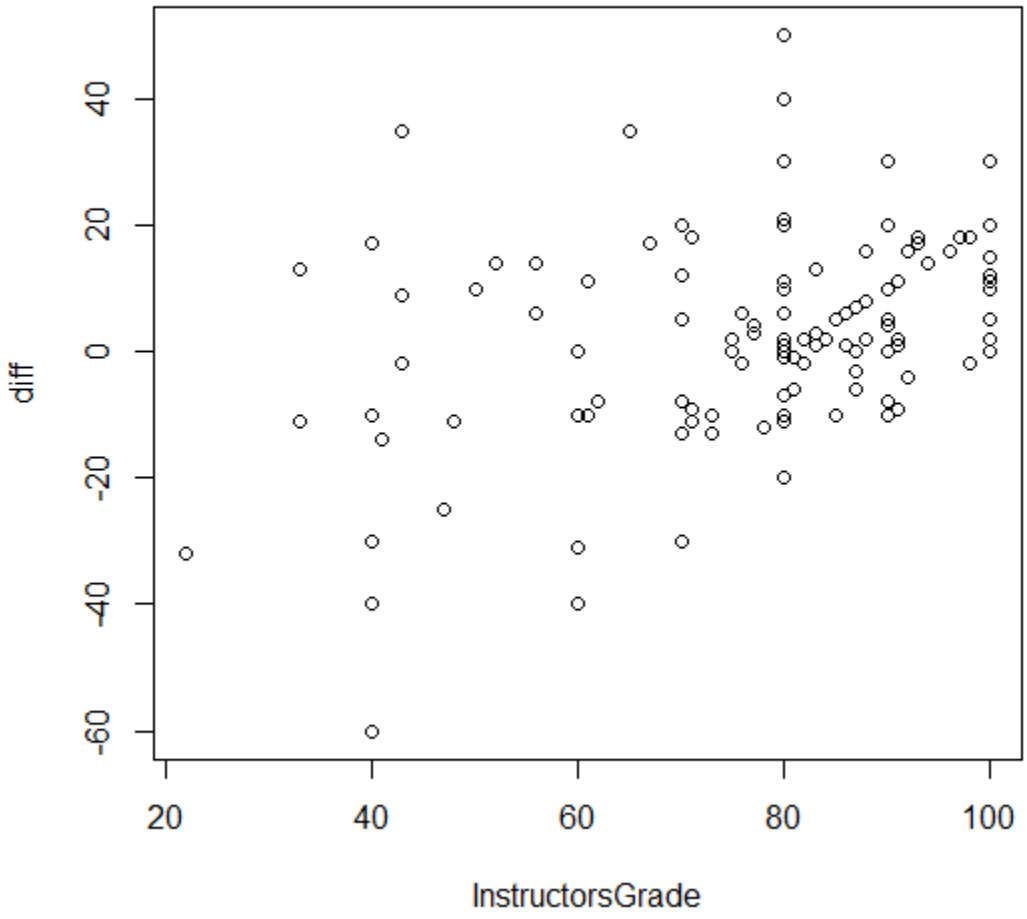
data: InstructorsGrade and StudentsGrade

t = 1.2133, df = 308.9, p-value = 0.2259

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1.55624 6.56261



Math 1120 Calculus I

	StudentsGrade	InstructorsGrade	diff
median	85.000000	81.0000000	-5.000000
mean	82.458824	75.7294118	-6.729412
SE.mean	1.659499	2.1109550	1.409093
CI.mean.0.95	3.300095	4.1978656	2.802136
var	234.084594	378.7711485	168.771148
std.dev	15.299823	19.4620438	12.991195
coef.var	0.185545	0.2569945	-1.930510

Welch Two Sample t-test

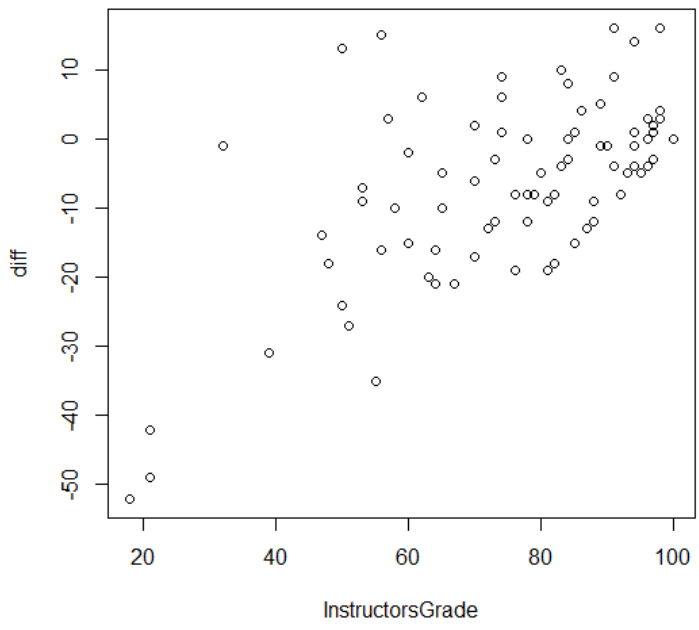
data: InstructorsGrade and StudentsGrade

t = -2.5062, df = 159.13, p-value = 0.01321

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-12.03255 -1.42627



Math191

	StudentsGrade	InstructorsGrade	diff
median	86.0000000	82.5000000	-6.500000
mean	86.9285714	80.7857143	-6.142857
SE.mean	2.4728510	3.3825320	3.520516
CI.mean.0.95	5.3422698	7.3075161	7.605613
var	85.6098901	160.1813187	173.516484
std.dev	9.2525613	12.6562759	13.172566
coef.var	0.1064387	0.1566648	-2.144371

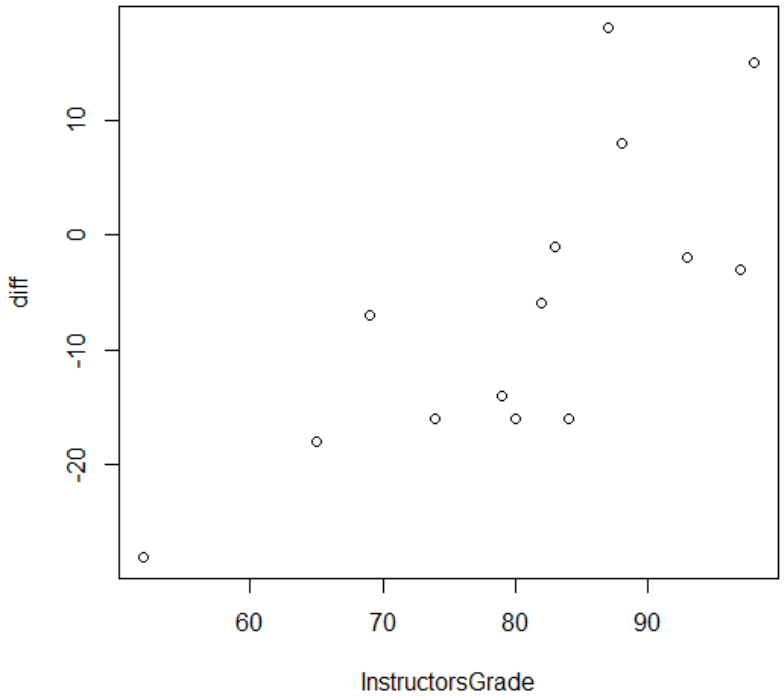
Welch Two Sample t-test

data: InstructorsGrade and StudentsGrade

t = -1.4661, df = 23.808, p-value = 0.1557

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: -14.794377 2.508663



Math192 Section 06

	StudentsGrade	InstructorsGrade	diff
median	90.0000000	77.000000	-4.000000
mean	83.5540070	73.790941	-9.763066
SE.mean	1.0742002	1.379446	1.264963
CI.mean.0.95	2.1143410	2.715154	2.489817
var	331.1710241	546.123974	459.237372
std.dev	18.1981050	23.369296	21.429824
coef.var	0.2178005	0.316696	-2.194989

Welch Two Sample t-test

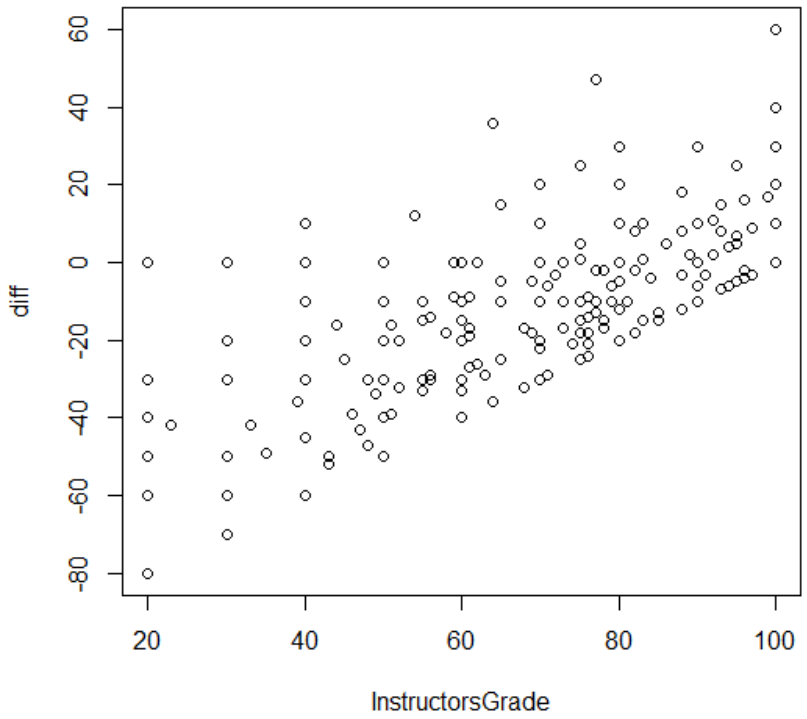
data: InstructorsGrade and StudentsGrade

t = -5.5841, df = 539.61, p-value = 3.727e-08

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-13.197500 -6.328632



Math 192 Section 8

	StudentsGrade	InstructorsGrade	diff
median	90.0000000	80.0000000	-7.000000
mean	88.1325301	78.2088353	-9.923695
SE.mean	0.8390781	1.2634247	1.076636
CI.mean.0.95	1.6526279	2.4884106	2.120516
var	175.3089778	397.4642765	288.627219
std.dev	13.2404297	19.9365061	16.989032
coef.var	0.1502332	0.2549137	-1.711966

Welch Two Sample t-test

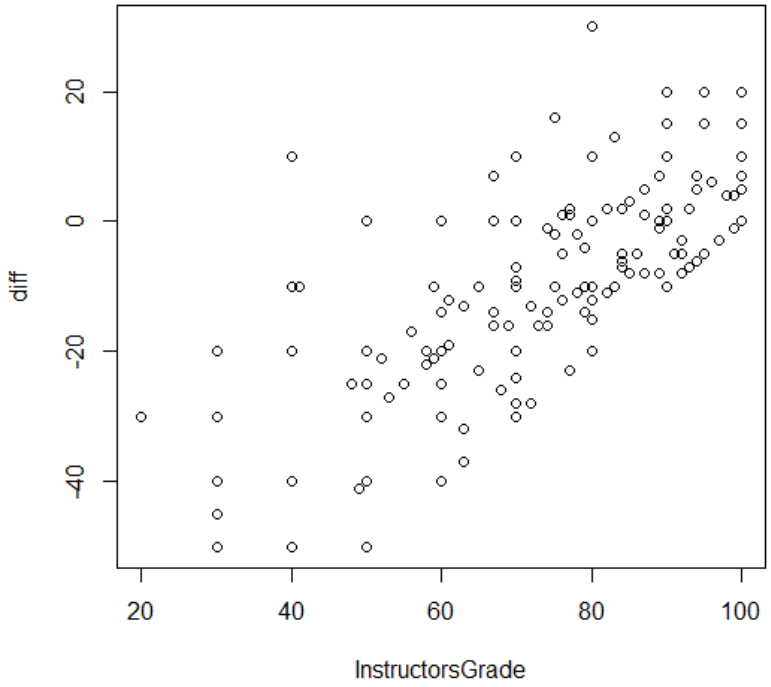
data: InstructorsGrade and StudentsGrade

t = -6.5431, df = 431.14, p-value = 1.712e-10

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-12.904686 -6.942704



Math2204 Elementary Statistics

	StudentsGrade	InstructorsGrade	diff
median	90.0000000	80.0000000	-10.000000
mean	85.5863874	74.2774869	-11.308901
SE.mean	0.9600210	1.4609484	1.241591
CI.mean.0.95	1.8936684	2.8817618	2.449073
var	176.0332874	407.6647010	294.435657
std.dev	13.2677537	20.1907083	17.159128
coef.var	0.1550218	0.2718281	-1.517312

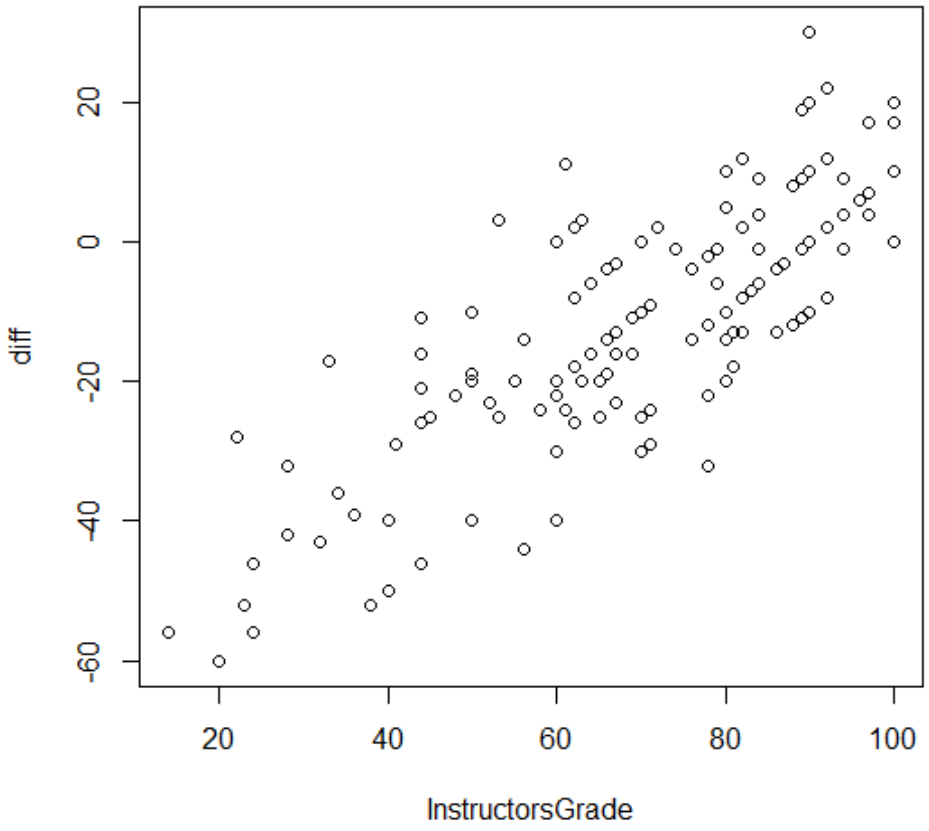
Welch Two Sample t-test

data: InstructorsGrade and StudentsGrade

t = -6.4691, df = 328.3, p-value = 3.58e-10

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: -14.747879 -7.869922



Math 2222/2470 Calculus III

	StudentsGrade	InstructorsGrade	diff
median	84.0000000	87.0000000	2.000000
mean	78.7623762	82.4851485	3.722772
SE.mean	1.6920401	1.6008422	1.427531
CI.mean.0.95	3.3569594	3.1760254	2.832181
var	289.1629703	258.8322772	205.822376
std.dev	17.0047926	16.0882652	14.346511

coef.var	0.2158999	0.1950444	3.853717

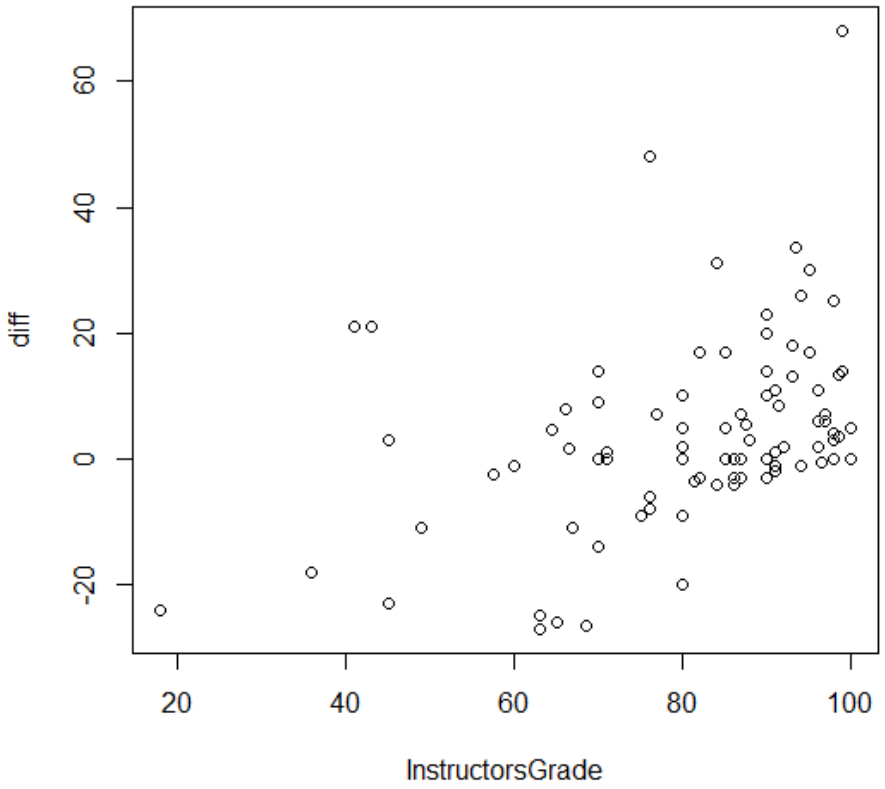
Welch Two Sample t-test

data: InstructorsGrade and StudentsGrade

t = 1.5982, df = 199.39, p-value = 0.1116

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: -0.8704754 8.3160199



Math3313 Differential Equations

	StudentsGrade	InstructorsGrade	diff
median	90.5000000	93.500000	2.500000
mean	86.2500000	90.083333	3.833333
SE.mean	4.6647047	3.122398	2.519119
CI.mean.0.95	10.2669458	6.872351	5.544543
var	261.1136364	116.992424	76.151515

std.dev	16.1590110	10.816304	8.726484
coef.var	0.1873509	0.120070	2.276474

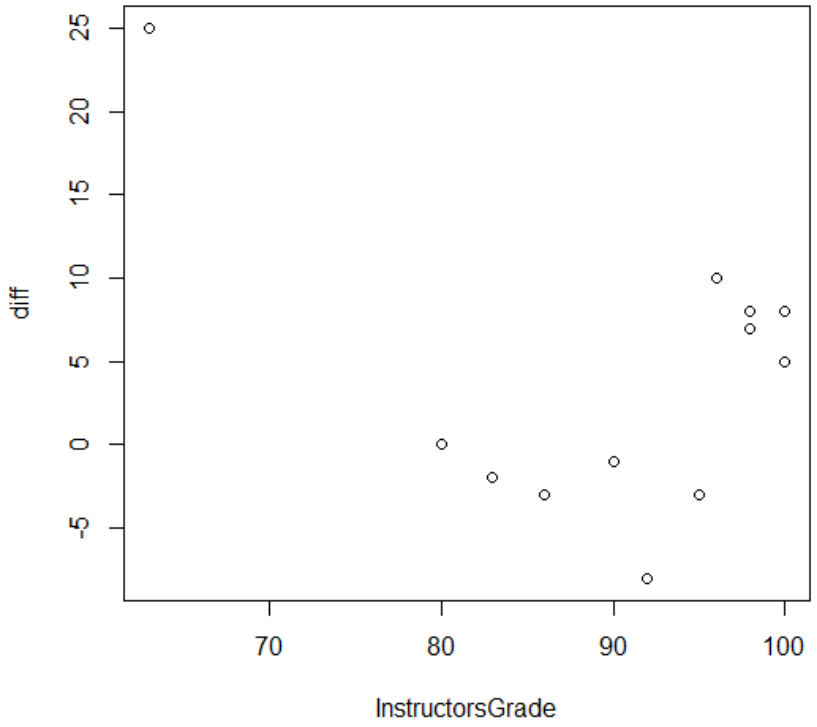
Welch Two Sample t-test

data: InstructorsGrade and StudentsGrade

t = 0.68291, df = 19.209, p-value = 0.5028

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: -7.906731 15.573398



Longitudinal Analysis

	diff1	diff2	diff3	diff4	avins	avdif
median	0.000000	5.0000000	-6.000000	-3.000000	78.0000000	-1.000000
mean	-5.576923	0.7362637	-5.725275	-6.340659	76.9450549	-4.208791
SE.mean	2.290221	1.8630679	1.963592	1.550550	1.3542645	1.374859
CI.mean.0.95	4.549923	3.7013096	3.901018	3.080438	2.6904828	2.731396

var	477.305128	315.8630037	350.868132	218.782662	166.8969475	172.011477
std.dev	21.847314	17.7725351	18.731474	14.791304	12.9188601	13.115315
coef.var	-3.917449	24.1388163	-3.271716	-2.332771	0.1678972	-3.116171

Relational Study between End of Semester Grades and Average Difference

Pearson's product-moment correlation

t = 8.5524, df = 89, p-value = 3.167e-13

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval: (0.5404526, 0.7709479)

sample estimates:

cor =0.6716414

Regression Analysis

Linear model
Average Difference= -56.67 + 0.6819*Average Semester Grades

Residuals:

Min	1Q	Median	3Q	Max
-17.9637	-7.5333	-0.6472	6.1258	26.3536

Coefficients:

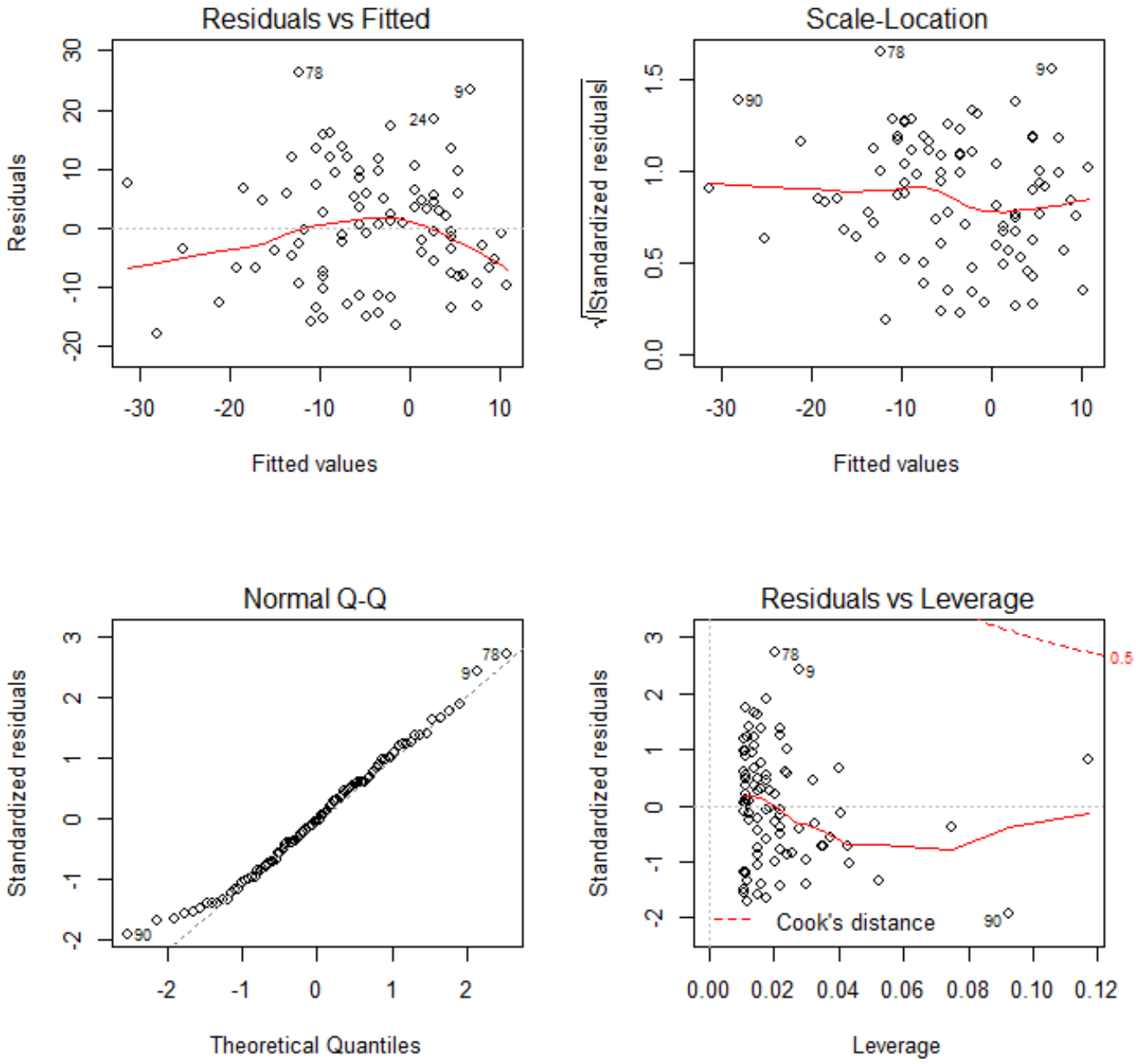
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-56.67416	6.21952	-9.112	2.20e-14
Ave. Sem. Grade	0.68185	0.07973	8.552	3.17e-13

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

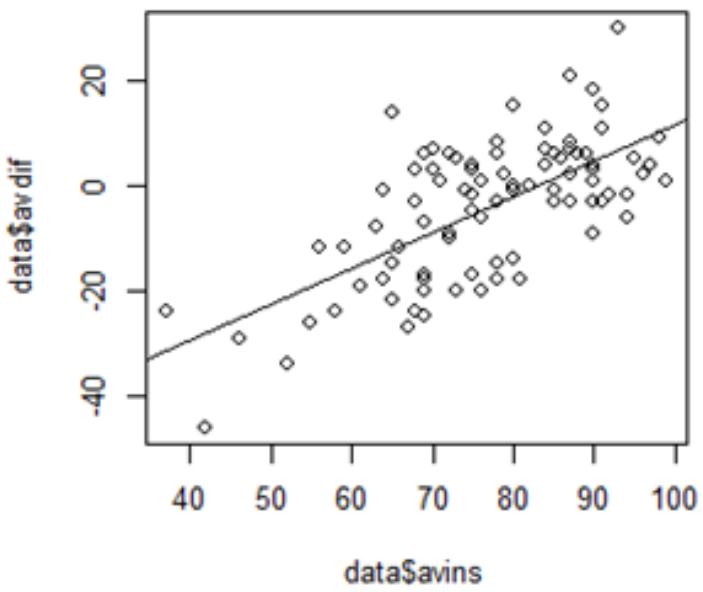
Residual standard error: 9.771 on 89 degrees of freedom

Multiple R-squared: 0.4511, Adjusted R-squared: 0.4449

F-statistic: 73.14 on 1 and 89 DF, p-value: 3.167e-13



Regression Line and Scatter plot (average semester grades versus average differences)



Beginning/End Comparison

	Beg. diff	End diff.	Ave. Ins. Grade	Ave. Diff.	Diff. of diff.
median	0.000000	-3.000000	78.0000000	-1.000000	-2.0000000
mean	-4.877358	-4.669811	77.1981132	-3.207547	0.2075472
SE.mean	2.243758	1.494267	1.2547427	1.301648	2.3439899
CI.mean.0.95	4.448957	2.962856	2.4879229	2.580928	4.6476990
var	533.651482	236.680413	166.8841869	179.594609	582.3946092
std.dev	23.100898	15.384421	12.9183663	13.401291	24.1328533
coef.var	-4.736354	-3.294442	0.1673404	-4.178050	116.2764751

Paired t-test

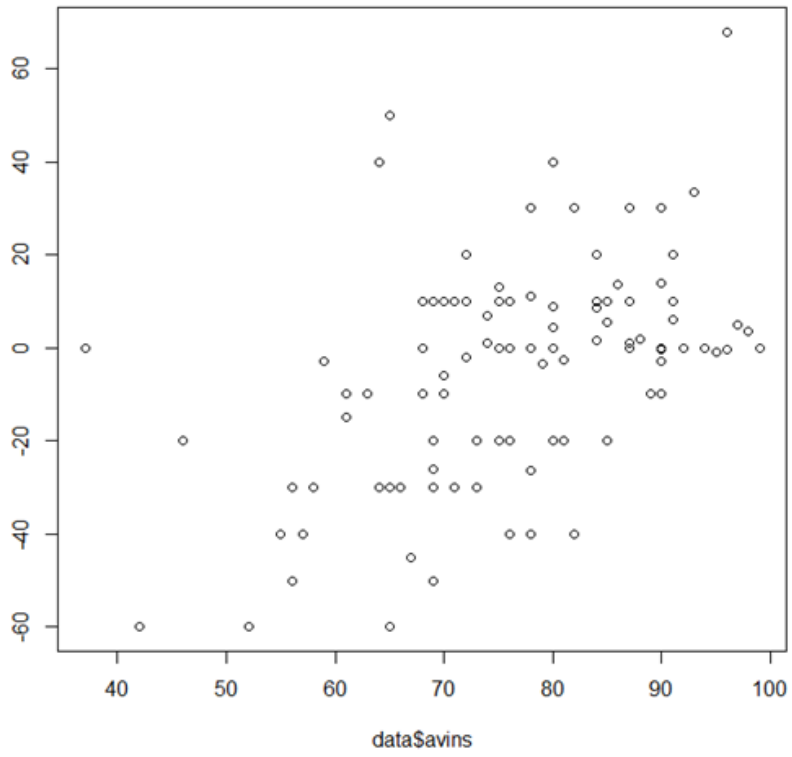
data: diff1 and diff4

t = 0.088544, df = 105, p-value = 0.9296

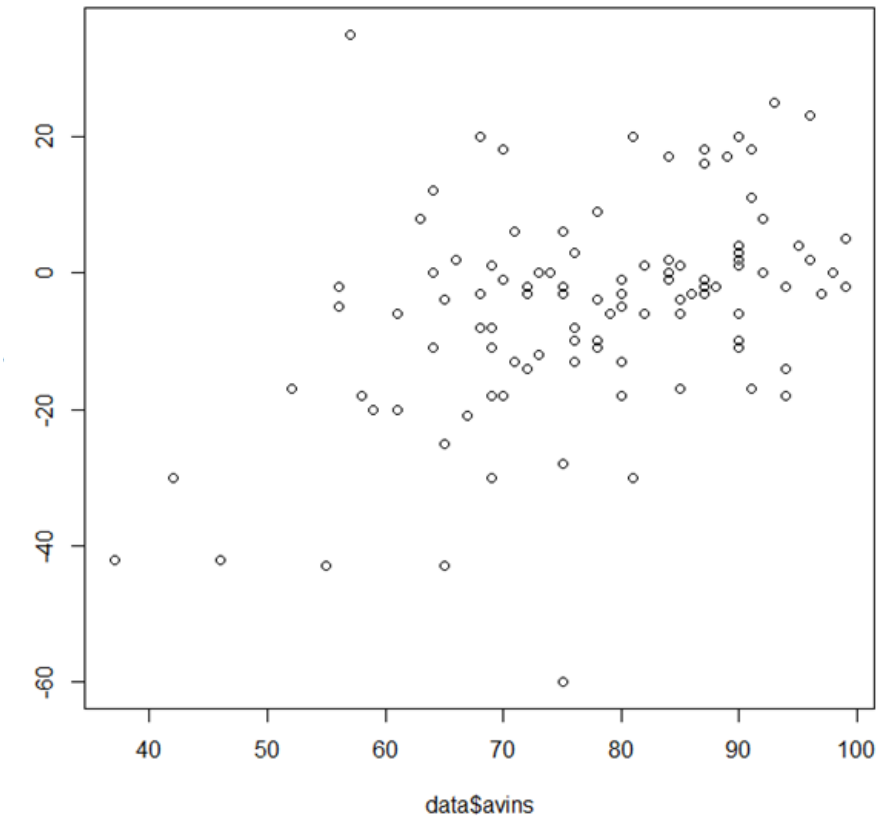
alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: (-4.440152 4.855246)

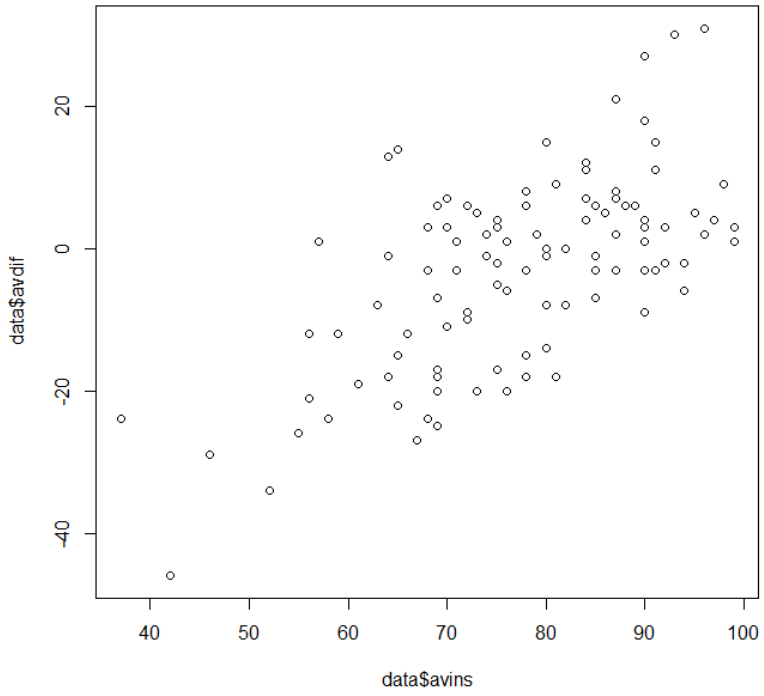
Scatter plot (Average Instructor's Grade versus beginning difference)



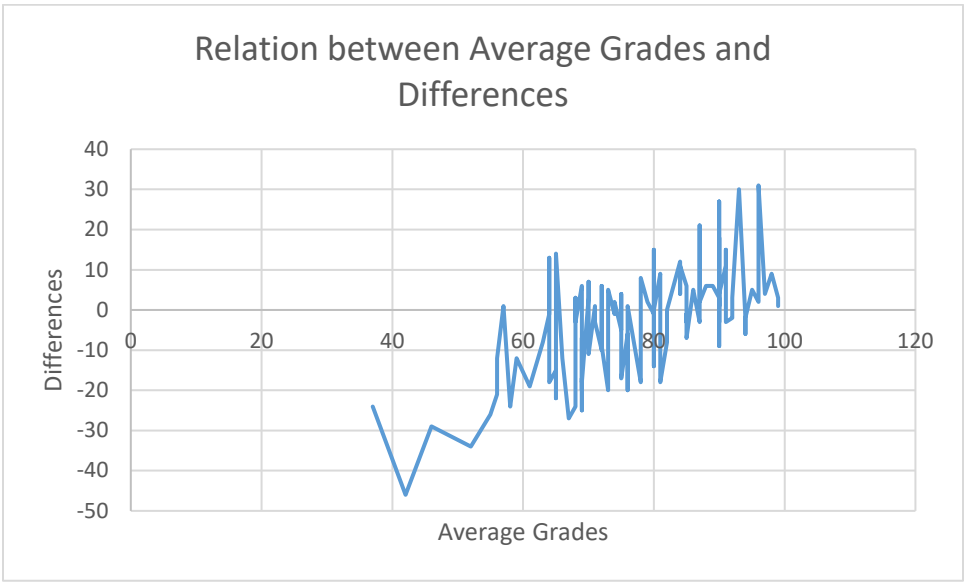
Scatter plot (average instructor's grade versus end difference)



Scatter plot (average instructor's grade versus average difference)



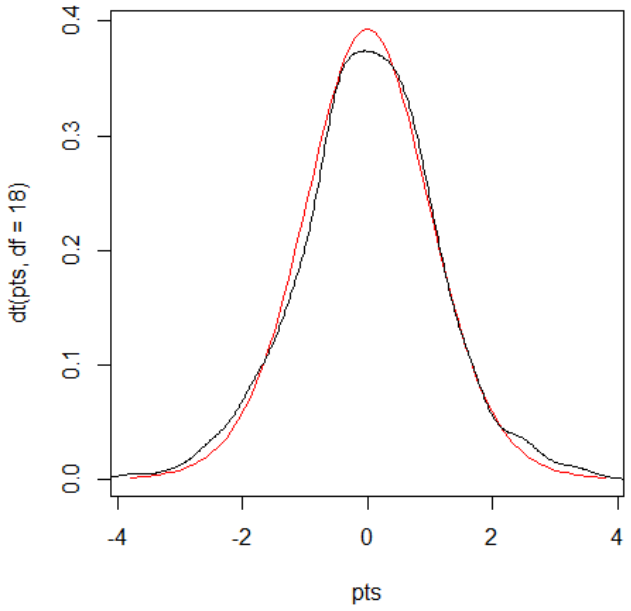
Line graph (Average Instructor's Grade versus Average difference)



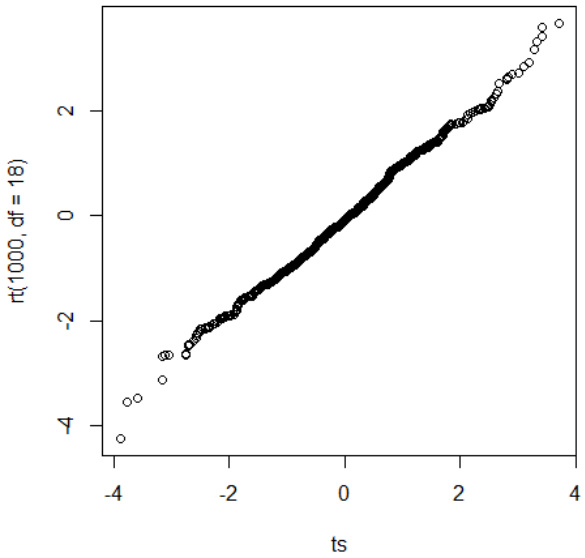
This clearly shows that students with high average grades underestimate while the students with low average grades over estimate.

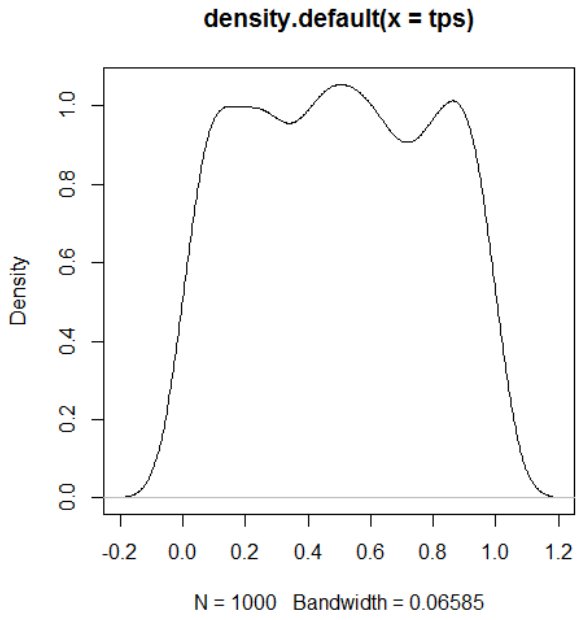
Error Analysis

Density Plot

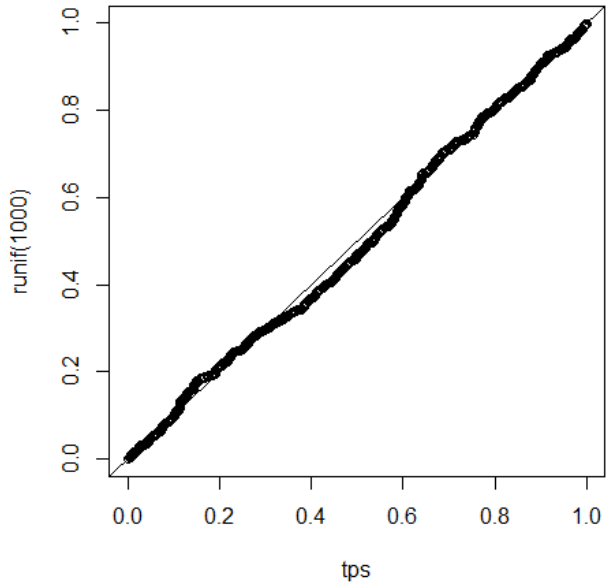


Quantile-Quantile Plot




density.default(x = tps)
Density
0.0
0.2
0.4
0.6
0.8
1.0
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
N = 1000 Bandwidth = 0.06585

QQ plot


runif(1000)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
tps

Math 320

	studentsGrades	instructorsGrade	diff
nbr.val	63.0000000	63.0000000	63.000000
nbr.null	0.0000000	0.0000000	5.000000
nbr.na	0.0000000	0.0000000	0.000000
min	7.0000000	20.0000000	-33.000000
max	100.0000000	100.0000000	20.000000
range	93.0000000	80.0000000	53.000000
sum	5205.0000000	4831.0000000	-374.000000
median	85.0000000	76.0000000	-3.000000
mean	82.6190476	76.6825397	-5.936508
SE.mean	1.9663618	2.1940927	1.528844
CI.mean.0.95	3.9307013	4.3859288	3.056116
var	243.5944700	303.2846902	147.253968
std.dev	15.6075133	17.4150708	12.134825
coef.var	0.1889094	0.2271061	-2.044101

Paired t-test

data: x and y

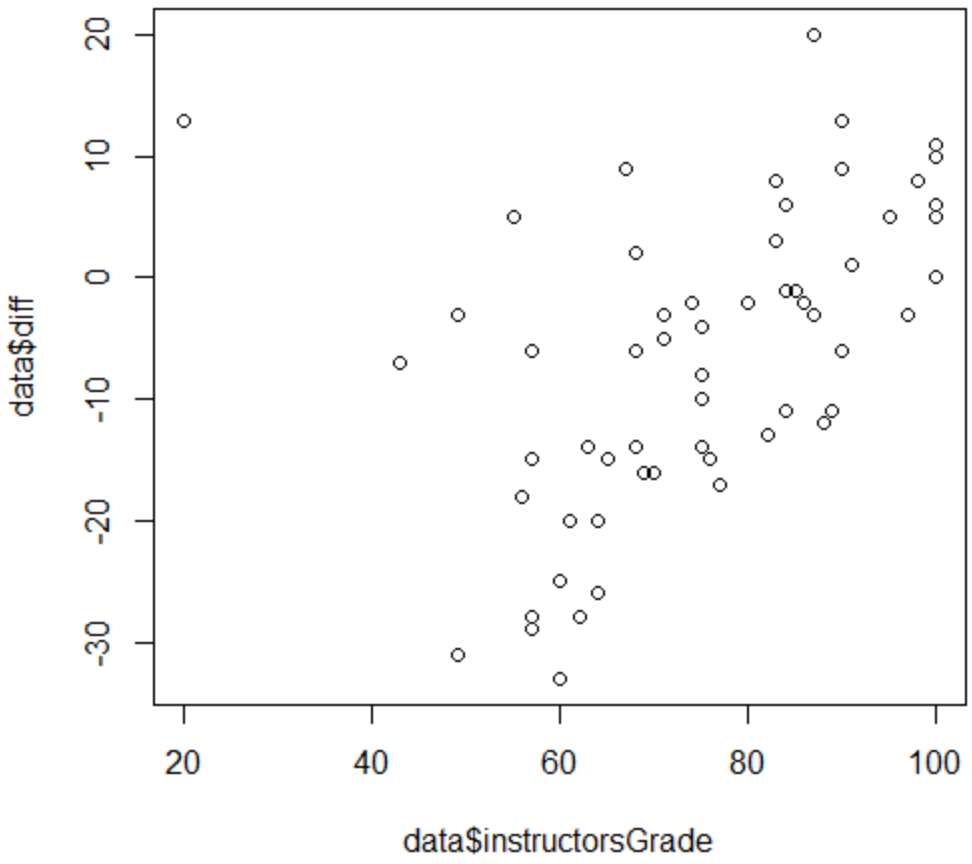
t = 3.883, df = 62, p-value = 0.000253

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval: 2.880392 8.992624

sample estimates: mean of the differences

5.936508


data$diff
20
10
0
-10
-20
-30
20
40
60
80
100
data$instructorsGrade