STA 441s20 Formulas and Printouts

Formula sheet is followed by

1. **Air Quality** SAS program and results
2. **Self-Esteem** SAS program and results
3. **UCLA graduate admissions** SAS program and results

$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1} x_{p-1} + \epsilon$ $\qquad SST = SSR + SSE$ $\qquad R^2 = \frac{SSR}{SST}$

$a = \frac{R_F^2 - R_R^2}{1 - R_R^2}$ $\qquad a = \frac{sF}{n-p+sF}$ $\qquad F = \left(\frac{n-p}{s}\right)\left(\frac{a}{1-a}\right)$

If an overall test has s numerator degrees of freedom and critical value c, the critical value of a Scheffé follow-up test with r degrees of freedom is $\left(\frac{s}{r}\right) \cdot c$.

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_{p-1} x_{p-1} = L \qquad \pi = \frac{e^L}{1+e^L}$$

$$\ln\left(\frac{\pi_1}{\pi_3}\right) = \beta_{0,1} + \beta_{1,1} x_1 + \ldots + \beta_{p-1,1} x_{p-1} = L_1 \qquad \pi_1 = \frac{e^{L_1}}{1+e^{L_1}+e^{L_2}}$$

$$\ln\left(\frac{\pi_2}{\pi_3}\right) = \beta_{0,2} + \beta_{1,2} x_1 + \ldots + \beta_{p-1,2} x_{p-1} = L_2 \qquad \pi_2 = \frac{e^{L_2}}{1+e^{L_1}+e^{L_2}}$$

$$\pi_3 = \frac{1}{1+e^{L_1}+e^{L_2}}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} E[y_1|\mathbf{X}=\mathbf{x}] \\ E[y_2|\mathbf{X}=\mathbf{x}] \\ \vdots \\ E[y_k|\mathbf{X}=\mathbf{x}] \end{pmatrix} = \begin{pmatrix} \beta_{0,1} + \beta_{1,1} x_1 + & \cdots & + \beta_{p-1,1} x_{p-1} \\ \beta_{0,2} + \beta_{1,2} x_1 + & \cdots & + \beta_{p-1,2} x_{p-1} \\ \vdots & \vdots & \vdots \\ \beta_{0,k} + \beta_{1,k} x_1 + & \cdots & + \beta_{p-1,k} x_{p-1} \end{pmatrix}$$

Unknown (`type=un`)

$$\begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \sigma_{3,4} \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 \end{pmatrix}$$

Compound Symmetry (`type=cs`)

$$\begin{pmatrix} \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 \end{pmatrix}$$

Autoregressive (`type=ar(1)`)

$$\sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

```
/******************** FinalAirQuality.sas ***************************/
title 'Air Quality Time Series';

data ozone;
     infile '/home/brunner0/441s20/airquality.data.txt' firstobs=2;
     input   Time Ozone SolarRad Wind Temp Month Day;


proc autoreg;
     title2 'Ozone with no explanatory variables';
     model ozone = /  dwprob;
run;

proc autoreg;
     title2 'Ozone as a function of SolarRad Wind Temp Time';
     model ozone = SolarRad Wind Temp Time /  dwprob;
run;
```

Air Quality Time Series
Ozone with no explanatory variables

The AUTOREG Procedure

Dependent Variable	Ozone

Air Quality Time Series
Ozone with no explanatory variables

The AUTOREG Procedure

Ordinary Least Squares Estimates			
SSE	125143.06	DFE	115
MSE	1088	Root MSE	32.98788
SBC	1144.04756	AIC	1141.29397
MAE	26.3501784	AICC	1141.32906
MAPE	149.067135	HQC	1142.41177
Durbin-Watson	1.0683	Total R-Square	0.0000

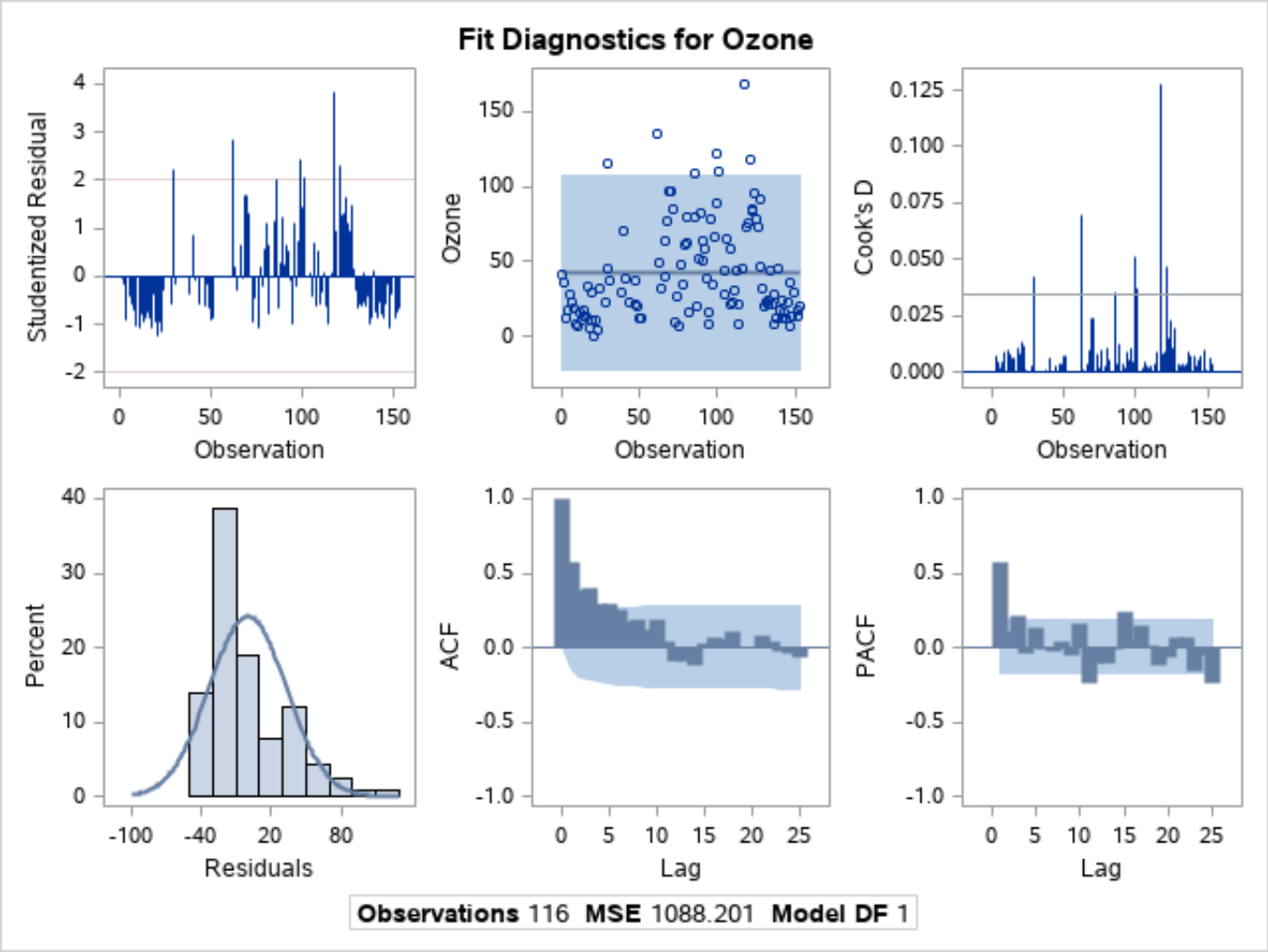
Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	1.0683	<.0001	1.0000

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	42.1293	3.0628	13.75	<.0001

Air Quality Time Series
Ozone with no explanatory variables

The AUTOREG Procedure



Air Quality Time Series
Ozone as a function of SolarRad Wind Temp Time

The AUTOREG Procedure

Dependent Variable	Ozone

Air Quality Time Series
Ozone as a function of SolarRad Wind Temp Time

The AUTOREG Procedure

Ordinary Least Squares Estimates			
SSE	46759.6175	DFE	106
MSE	441.12847	Root MSE	21.00306
SBC	1009.3522	AIC	995.804553
MAE	15.1959742	AICC	996.375982
MAPE	65.93552	HQC	1001.30043
Durbin-Watson	2.0004	Total R-Square	0.6161

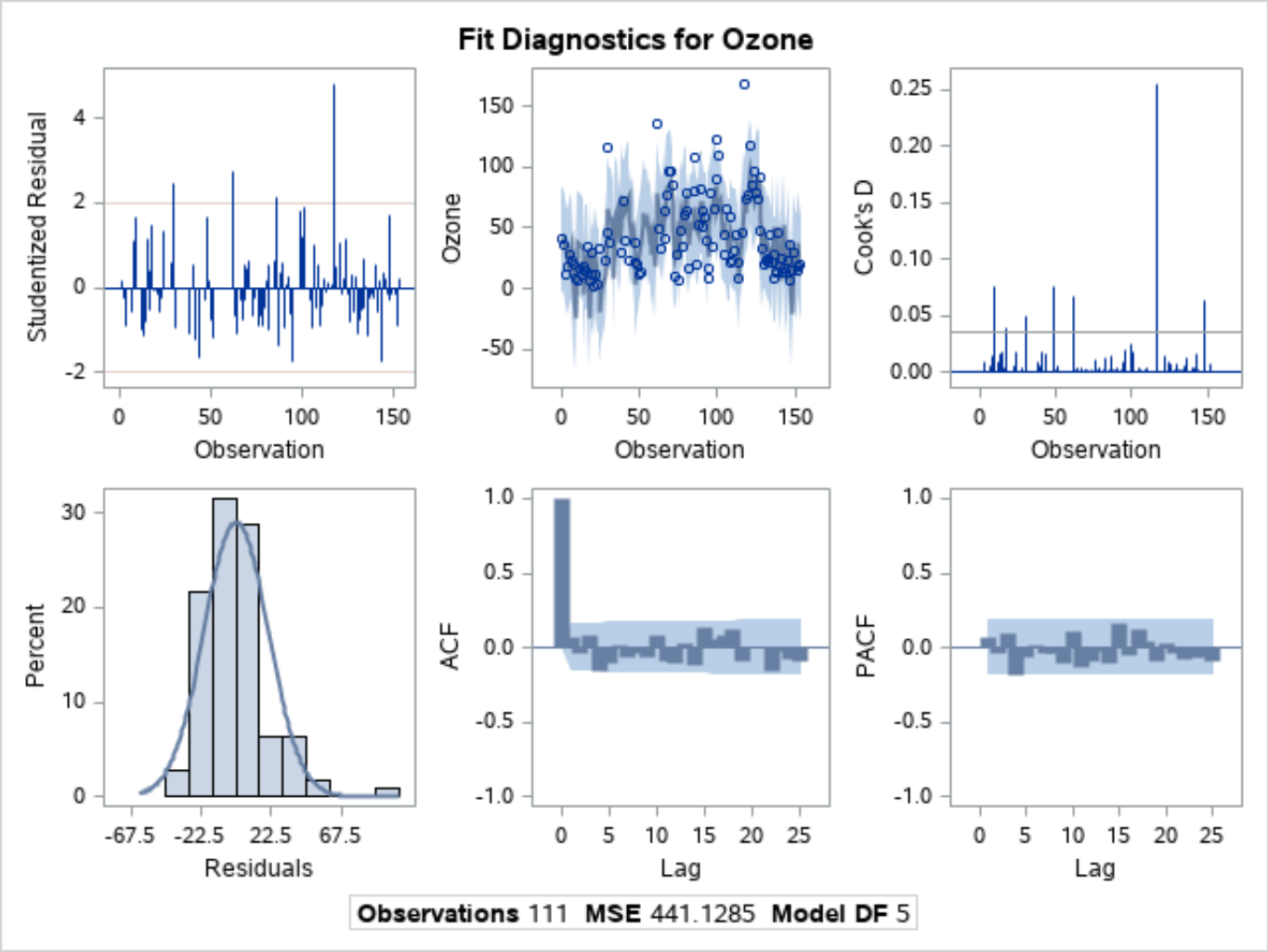
Durbin-Watson Statistics			
Order	DW	Pr < DW	Pr > DW
1	2.0004	0.4298	0.5702

NOTE: Pr<DW is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Parameter Estimates					
Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-69.6860	23.0819	-3.02	0.0032
SolarRad	1	0.0511	0.0236	2.17	0.0325
Wind	1	-3.3181	0.6490	-5.11	<.0001
Temp	1	1.8268	0.2721	6.71	<.0001
Time	1	-0.0809	0.0482	-1.68	0.0961

Air Quality Time Series
Ozone as a function of SolarRad Wind Temp Time

The AUTOREG Procedure


Fit Diagnostics for Ozone
Studentized Residual
Observation
Ozone
Observation
Cook's D
Observation
Percent
Residuals
ACF
Lag
PACF
Lag
Observations 111 MSE 441.1285 Model DF 5

```
/******************** FinalSelfEsteem.sas **************************/
title 'Self-Esteem Data for Final Exam';

data psych; /* Multivariate data read */
     infile '/home/brunner0/441s20/selfesteem3.data.txt' firstobs=2;
     input   line1 id1 treatment1 $ Control1 Control2 Control3
             line2 id2 treatment2 $ Diet1    Diet2    Diet3;
     A   = Diet1+Diet2+Diet3 - (Control1+Control2+Control3);
     B1  = Diet1+Control1 - (Diet2+Control2);
     B2  = Diet2+Control2 - (Diet3+Control3);
     AB1 = Diet1-Control1 - (Diet2-Control2);
     AB2 = Diet2-Control2 - (Diet3-Control3);
     Diff1 = Diet1-Control1;
     Diff2 = Diet2-Control2;
     Diff3 = Diet3-Control3;

data utility; /* Univariate data read */
     infile '/home/brunner0/441s20/selfesteem3.data.txt' firstobs=2;
     input line id Treatment $ t1 t2 t3;
     /* But t1-t3 are still on the same line. */

data psych0;
     set utility;
     se = t1; Time = 1; output;
     se = t2; Time = 2; output;
     se = t3; Time = 3; output;
     label se = 'Self-Esteem';
     keep Treatment Time id se;

/* proc glm to get means and interaction plot */
proc glm data=psych0;
     title2 'Two-way Between';
     class Time Treatment;
     model se = Time|Treatment;
     means Time*Treatment;
run;

proc means t probt data=psych;
     title2 'Matched t-tests with proc means';
     var A -- Diff3;
run;

/* Multivariate tests for multiple contrasts */

quit; /* Need quit to make ods select work. */
ods select MultStat;
proc reg data = psych;
     title2 'Test E(B1) = E(B2) = 0';
     model B1 B2 = ;
     mtest intercept=0;
run;

quit;
ods select MultStat;
proc reg data = psych;
     title2 'Test E(AB1) = E(AB2) = 0';
     model AB1 AB2 = ;
     mtest intercept=0;
run;

quit;
ods select MultStat;
proc reg data = psych;
     title2 'Test E(Diff1) = E(Diff2) = E(Diff3) 0';
     model Diff1 Diff2 Diff3 = ;
     mtest intercept=0;
run;
```

Self-Esteem Data for Final Exam
Two-way Between

The GLM Procedure

Class Level Information		
Class	Levels	Values
Time	3	1 2 3
Treatment	2	Diet ctr

Number of Observations Read	72
Number of Observations Used	72

Self-Esteem Data for Final Exam
Two-way Between

The GLM Procedure

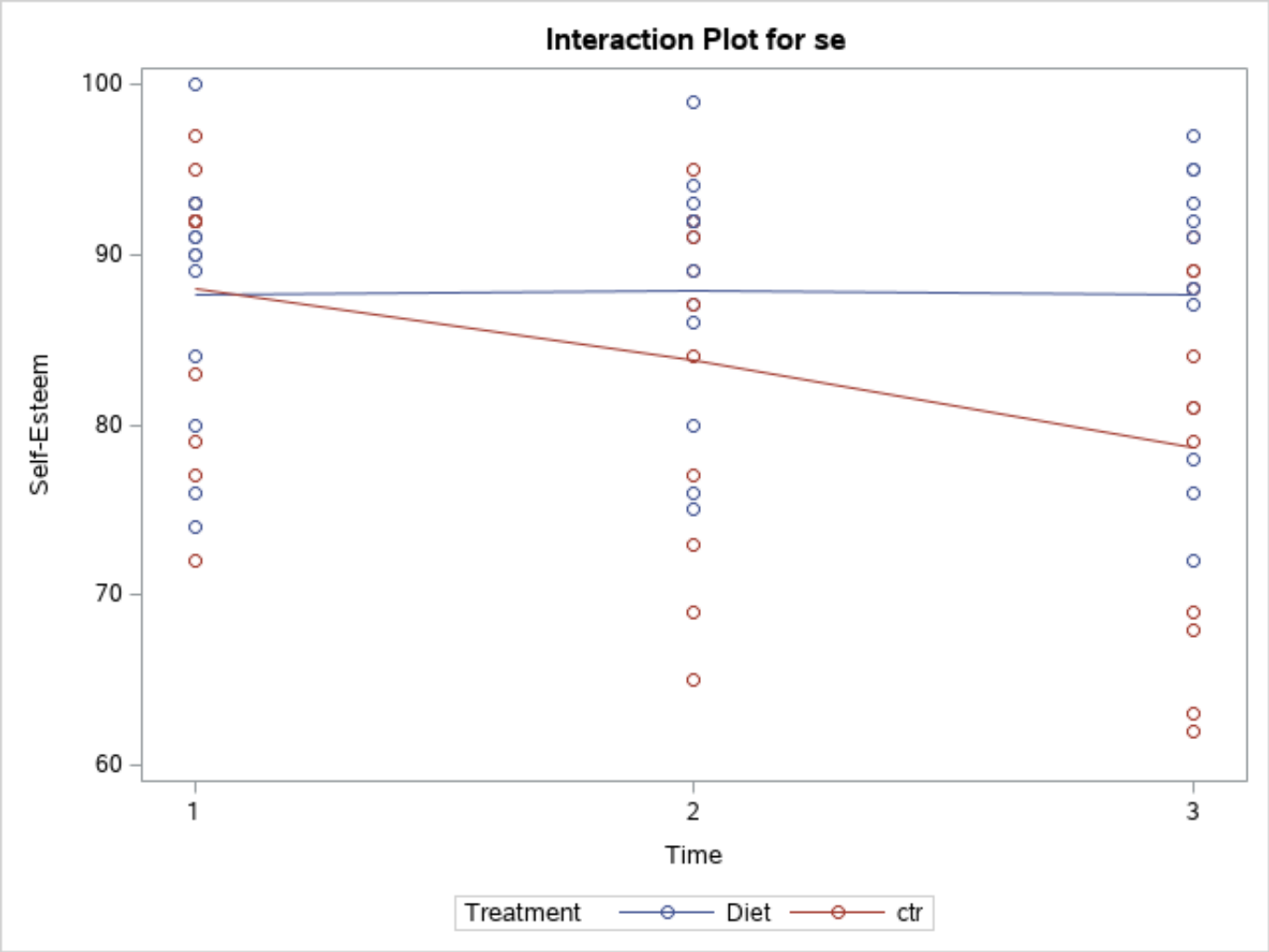
Dependent Variable: se Self-Esteem

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	841.736111	168.347222	2.19	0.0654
Error	66	5065.583333	76.751263		
Corrected Total	71	5907.319444			

R-Square	Coeff Var	Root MSE	se Mean
0.142490	10.23489	8.760780	85.59722

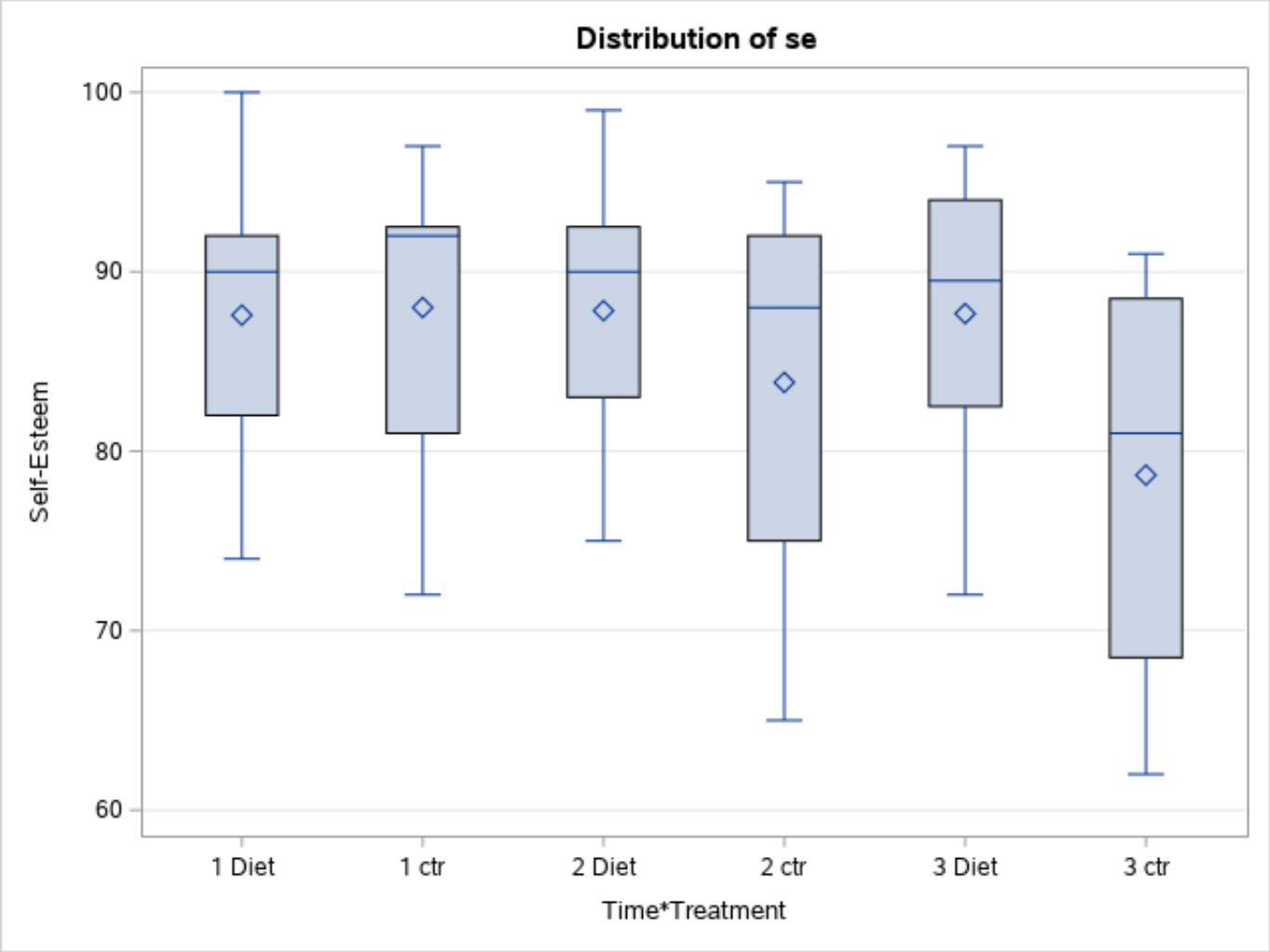
Source	DF	Type I SS	Mean Square	F Value	Pr > F
Time	2	258.6944444	129.3472222	1.69	0.1933
Treatment	1	316.6805556	316.6805556	4.13	0.0463
Time*Treatment	2	266.3611111	133.1805556	1.74	0.1843

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Time	2	258.6944444	129.3472222	1.69	0.1933
Treatment	1	316.6805556	316.6805556	4.13	0.0463
Time*Treatment	2	266.3611111	133.1805556	1.74	0.1843



Self-Esteem Data for Final Exam
Two-way Between

The GLM Procedure



Level of Time	Level of Treatment	N	se	
			Mean	Std Dev
1	Diet	12	87.5833333	7.6212422
1	ctr	12	88.0000000	8.0791539
2	Diet	12	87.8333333	7.4202834
2	ctr	12	83.8333333	10.2321537
3	Diet	12	87.6666667	8.1389449
3	ctr	12	78.6666667	10.5428419

Self-Esteem Data for Final Exam
Matched t-tests with proc means

The MEANS Procedure

Variable	t Value	Pr > |t|
A	3.94	0.0023
B1	4.35	0.0012
B2	4.83	0.0005
AB1	-3.82	0.0029

AB2	-5.80	0.0001
Diff1	-0.61	0.5521
Diff2	3.00	0.0120
Diff3	5.56	0.0002

Self-Esteem Data for Final Exam
Test E(B1) = E(B2) = 0

The REG Procedure
Model: MODEL1
Multivariate Test 1

Multivariate Statistics and Exact F Statistics					
S=1 M=0 N=4					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.25759012	14.41	2	10	0.0011
Pillai's Trace	0.74240988	14.41	2	10	0.0011
Hotelling-Lawley Trace	2.88213640	14.41	2	10	0.0011
Roy's Greatest Root	2.88213640	14.41	2	10	0.0011

Self-Esteem Data for Final Exam
Test E(AB1) = E(AB2) = 0

The REG Procedure
Model: MODEL1
Multivariate Test 1

Multivariate Statistics and Exact F Statistics					
S=1 M=0 N=4					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.19993507	20.01	2	10	0.0003
Pillai's Trace	0.80006493	20.01	2	10	0.0003
Hotelling-Lawley Trace	4.00162371	20.01	2	10	0.0003
Roy's Greatest Root	4.00162371	20.01	2	10	0.0003

Self-Esteem Data for Final Exam
Test E(Diff1) = E(Diff2) = E(Diff3) 0

The REG Procedure
Model: MODEL1
Multivariate Test 1

Multivariate Statistics and Exact F Statistics					
S=1 M=0.5 N=3.5					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.19992229	12.01	3	9	0.0017

Pillai's Trace	0.80007771	12.01	3	9	0.0017
Hotelling-Lawley Trace	4.00194345	12.01	3	9	0.0017
Roy's Greatest Root	4.00194345	12.01	3	9	0.0017

```
/********************** FinalUCLA.sas ************************/
title 'UCLA Grad School Data';

data ucla;
     infile '/home/brunner0/441s20/ucla.data.txt' firstobs=2;
     input id admit gre gpa rank;
     if rank=2 then r2=1; else r2=0;
     if rank=3 then r3=1; else r3=0;
     if rank=4 then r4=1; else r4=0;

proc logistic data=ucla;
     model admit (event='1') = gre gpa r2 r3 r4;
     prestige: test r2=r3=r4=0;
```

UCLA Grad School Data

The LOGISTIC Procedure

Model Information	
Data Set	WORK.UCLA
Response Variable	admit
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	400
Number of Observations Used	400

Response Profile		
Ordered Value	admit	Total Frequency
1	0	273
2	1	127

Probability modeled is admit=1.

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	501.977	470.517
SC	505.968	494.466
-2 Log L	499.977	458.517

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	41.4590	5	<.0001
Score	40.1603	5	<.0001
Wald	36.1390	5	<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.9899	1.1399	12.2507	0.0005
gre	1	0.00226	0.00109	4.2842	0.0385
gpa	1	0.8040	0.3318	5.8714	0.0154
r2	1	-0.6754	0.3165	4.5547	0.0328
r3	1	-1.3402	0.3453	15.0636	0.0001
r4	1	-1.5514	0.4178	13.7870	0.0002

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
gre	1.002	1.000	1.004
gpa	2.235	1.166	4.282
r2	0.509	0.274	0.946
r3	0.262	0.133	0.515
r4	0.212	0.093	0.481

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	69.3	Somers' D	0.386
Percent Discordant	30.7	Gamma	0.386
Percent Tied	0.0	Tau-a	0.168
Pairs	34671	c	0.693

Linear Hypotheses Testing Results			
Label	Wald Chi-Square	DF	Pr > ChiSq
prestige	20.8949	3	0.0001