# STA 441s16 Formulas and Printouts

## Formulas

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

$a = \frac{R^2_F - R^2_R}{1 - R^2_R}$ $\quad a = \frac{sF}{n-p+sF}$ $\quad F = \left(\frac{n-p}{s}\right)\left(\frac{a}{1-a}\right)$

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 $\left(\frac{s}{r}\right) \cdot c$.

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

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

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

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

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)`)

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

---

## Printouts

1. The TV data
2. The Donner Party Data
3. The Math Data
4. Beat the Blues Data

# The TV Data

```
/* FinalTV.sas */
title 'The TV Data';

proc format;
     value locfmt  1 = 'Rural'
                   2 = 'Small Town'
                   3 = 'City';
     value ynfmt  0 = 'No'  1 = 'Yes';
data tv;
     infile '/folders/myfolders/TV1.data.txt' firstobs=2; /* Skip the header */
     input dist hsehold value q1-q9;
     label value = 'Value of house in $US'
           q1 = 'Numbr persons 12 and older in household'
           q2 = 'Numbr persons 11 and younger in household'
           q3 = 'Numbr TV sets in Household'
           q4 = 'Price willing to pay per month for cable'
           q5 = 'Total TV hours watched last week'
           q6 = 'Hours Public Affairs watched last week'
           q7 = 'Hours Sports watched last week'
           q8 = 'Hours Children''s programming watched last week'
           q9 = 'Hours Movies watched last week';
     if 1 <= dist <= 25 then location=1;
        else if 26 <= dist <= 50 then location=2;
        else if 51 <= dist <= 75 then location=3;
        else location=.;
        format location locfmt.;
      /* Dummy variables for location */
      if location = 1 then L1 = 1;
         else if location = 2 then L1 = 0;
         else if location = 3 then L1 = 0;
      if location = 2 then L2 = 1;
         else if location = 1 then L2 = 0;
         else if location = 3 then L2 = 0;
       npeople = q2+q3;
       label npeople = 'Total Number of people in household';

proc means;
     var value q3-q5;

/* I will center the main continuous predictors, but first
   get rid of cases with any missing values on my variables. */

data tv2;
     set tv;
     utility = value+q3+q4+q5+L1+L2;
     if utility ne .; /* Otherwise delete */

proc standard out=tv3 mean=0;
     var value q3 q5;

proc reg data=tv3 plots=none;
     model q4 = L1 L2  value q3 q5;
     Test1: test q3=q5;
     Test2: test q3=q5=0;
     Test3: test L1=L2;
     Test4: test L1=L2=0;
```

# The TV Data

### The MEANS Procedure

| Variable | Label | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|---|
| value | Value of house in $US | 500 | 69368.26 | 16455.73 | 25087.00 | 124105.00 |
| q3 | Numbr TV sets in Household | 477 | 1.9014675 | 1.1360859 | 0 | 5.0000000 |
| q4 | Price willing to pay per month for cable | 456 | 10.3947368 | 6.6430080 | 0 | 25.0000000 |
| q5 | Total TV hours watched last week | 477 | 56.3018868 | 43.0521115 | 0 | 181.0000000 |

# The TV Data

### The REG Procedure
### Model: MODEL1
### Dependent Variable: q4 Price willing to pay per month for cable

| | |
|---|---|
| Number of Observations Read | 456 |
| Number of Observations Used | 456 |

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 5 | 14852 | 2970.30539 | 255.70 | <.0001 |
| Error | 450 | 5227.42042 | 11.61649 | | |
| Corrected Total | 455 | 20079 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 3.40830 | R-Square | 0.7397 |
| Dependent Mean | 10.39474 | Adj R-Sq | 0.7368 |
| Coeff Var | 32.78868 | | |

| Parameter Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
| Intercept | Intercept | 1 | 10.10151 | 0.21626 | 46.71 | <.0001 |
| L1 | | 1 | 1.08055 | 0.44074 | 2.45 | 0.0146 |
| L2 | | 1 | 0.44818 | 0.47458 | 0.94 | 0.3455 |
| value | Value of house in $US | 1 | 0.00024693 | 0.00001136 | 21.74 | <.0001 |
| q3 | Numbr TV sets in Household | 1 | 0.95334 | 0.16376 | 5.82 | <.0001 |
| q5 | Total TV hours watched last week | 1 | 0.06931 | 0.00467 | 14.85 | <.0001 |

# The TV Data

### The REG Procedure
### Model: MODEL1

| Test Test1 Results for Dependent Variable q4 | | | | |
|---|---|---|---|---|
| Source | DF | Mean Square | F Value | Pr > F |

| | | | | |
|---|---|---|---|---|
| **Numerator** | 1 | 329.94531 | 28.40 | <.0001 |
| **Denominator** | 450 | 11.61649 | | |

## The TV Data

**The REG Procedure**
**Model: MODEL1**

| Test Test2 Results for Dependent Variable q4 | | | | |
|---|---|---|---|---|
| **Source** | **DF** | **Mean Square** | **F Value** | **Pr > F** |
| **Numerator** | 2 | 2384.23233 | 205.25 | <.0001 |
| **Denominator** | 450 | 11.61649 | | |

## The TV Data

**The REG Procedure**
**Model: MODEL1**

| Test Test3 Results for Dependent Variable q4 | | | | |
|---|---|---|---|---|
| **Source** | **DF** | **Mean Square** | **F Value** | **Pr > F** |
| **Numerator** | 1 | 16.85721 | 1.45 | 0.2290 |
| **Denominator** | 450 | 11.61649 | | |

## The TV Data

**The REG Procedure**
**Model: MODEL1**

| Test Test4 Results for Dependent Variable q4 | | | | |
|---|---|---|---|---|
| **Source** | **DF** | **Mean Square** | **F Value** | **Pr > F** |
| **Numerator** | 2 | 34.97945 | 3.01 | 0.0502 |
| **Denominator** | 450 | 11.61649 | | |

# The Donner Party Data

```
/* FinalDonner.sas */
title 'The Donner Party Data';

proc format;
     value ynfmt  0 = 'No'  1 = 'Yes';
     value sexfmt 0 = 'Male' 1 = 'Female';
data donner;
     infile '/folders/myfolders/donner.data.txt' firstobs=2;
     input Age Sex Survival;
     agesex = Age*Sex;
     format Survival ynfmt.; format Sex sexfmt.;

proc logistic;
     model Survival (event='Yes') = Age Sex;
```

## The LOGISTIC Procedure

| Model Information | |
|---|---|
| **Data Set** | WORK.DONNER |
| **Response Variable** | Survival |
| **Number of Response Levels** | 2 |
| **Model** | binary logit |
| **Optimization Technique** | Fisher's scoring |

| | |
|---|---|
| **Number of Observations Read** | 45 |
| **Number of Observations Used** | 45 |

| Response Profile | | |
|---|---|---|
| **Ordered Value** | **Survival** | **Total Frequency** |
| **1** | No | 25 |
| **2** | Yes | 20 |

**Probability modeled is Survival='Yes'.**

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

| Model Fit Statistics | | |
|---|---|---|
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 63.827 | 57.256 |
| **SC** | 65.633 | 62.676 |
| **-2 Log L** | 61.827 | 51.256 |

| Testing Global Null Hypothesis: BETA=0 | | | |
|---|---|---|---|
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 10.5703 | 2 | 0.0051 |
| **Score** | 9.0965 | 2 | 0.0106 |
| **Wald** | 6.8627 | 2 | 0.0323 |

| Analysis of Maximum Likelihood Estimates | | | | | |
|---|---|---|---|---|---|
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **Wald Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | 1.6331 | 1.1102 | 2.1637 | 0.1413 |
| **Age** | 1 | -0.0782 | 0.0373 | 4.3988 | 0.0360 |
| **Sex** | 1 | 1.5973 | 0.7555 | 4.4699 | 0.0345 |

## The LOGISTIC Procedure

| Odds Ratio Estimates | | | |
|---|---|---|---|
| Effect | Point Estimate | 95% Wald Confidence Limits | |
| Age | 0.925 | 0.860 | 0.995 |
| Sex | 4.940 | 1.124 | 21.716 |

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 73.0 | Somers' D | 0.492 |
| Percent Discordant | 23.8 | Gamma | 0.508 |
| Percent Tied | 3.2 | Tau-a | 0.248 |
| Pairs | 500 | c | 0.746 |

# The Math Data

```
/* FinalMath.sas */
%include '/folders/myfolders/441s16/Lecture/readmath2b.sas';
/* Creates data table mathex */
title2 'The Math Data';

/*
proc freq;
     tables course2;
*/

proc logistic data = mathex;
     model course2 (ref='Mainstrm') = hsgpa hscalc hsengl
                                      / link = glogit;
```

# Prediction of Performance in First-year Calculus
# The Math Data

## The LOGISTIC Procedure

| Model Information | |
|---|---|
| **Data Set** | WORK.MATHEX |
| **Response Variable** | course2 |
| **Number of Response Levels** | 3 |
| **Model** | generalized logit |
| **Optimization Technique** | Newton-Raphson |

| | |
|---|---|
| **Number of Observations Read** | 579 |
| **Number of Observations Used** | 373 |

| Response Profile | | |
|---|---|---|
| **Ordered Value** | **course2** | **Total Frequency** |
| **1** | Catch-up | 21 |
| **2** | Elite | 30 |
| **3** | Mainstrm | 322 |

**Logits modeled use course2='Mainstrm' as the reference category.**

**Note:** 206 observations were deleted due to missing values for the response or explanatory variables.

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

| Model Fit Statistics | | |
|---|---|---|
| **Criterion** | **Intercept Only** | **Intercept and Covariates** |
| **AIC** | 370.745 | 333.709 |
| **SC** | 378.588 | 365.082 |
| **-2 Log L** | 366.745 | 317.709 |

| Testing Global Null Hypothesis: BETA=0 | | | |
|---|---|---|---|
| **Test** | **Chi-Square** | **DF** | **Pr > ChiSq** |
| **Likelihood Ratio** | 49.0356 | 6 | <.0001 |
| **Score** | 48.2763 | 6 | <.0001 |
| **Wald** | 38.4161 | 6 | <.0001 |

# Prediction of Performance in First-year Calculus
# The Math Data

## The LOGISTIC Procedure

| Type 3 Analysis of Effects | | | |
|---|---|---|---|
| Effect | DF | Wald Chi-Square | Pr > ChiSq |
| hsgpa | 2 | 0.3640 | 0.8336 |
| hscalc | 2 | 19.9470 | <.0001 |
| hsengl | 2 | 6.4923 | 0.0389 |

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | course2 | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
| Intercept | Catch-up | 1 | 10.5588 | 4.2647 | 6.1299 | 0.0133 |
| Intercept | Elite | 1 | -1.7202 | 2.6380 | 0.4252 | 0.5143 |
| hsgpa | Catch-up | 1 | -0.0407 | 0.0674 | 0.3635 | 0.5466 |
| hsgpa | Elite | 1 | 0.000050 | 0.0510 | 0.0000 | 0.9992 |
| hscalc | Catch-up | 1 | -0.1009 | 0.0253 | 15.9349 | <.0001 |
| hscalc | Elite | 1 | 0.0481 | 0.0260 | 3.4208 | 0.0644 |
| hsengl | Catch-up | 1 | -0.0435 | 0.0359 | 1.4675 | 0.2257 |
| hsengl | Elite | 1 | -0.0596 | 0.0259 | 5.2901 | 0.0214 |

| Odds Ratio Estimates | | | | |
|---|---|---|---|---|
| Effect | course2 | Point Estimate | 95% Wald Confidence Limits | |
| hsgpa | Catch-up | 0.960 | 0.841 | 1.096 |
| hsgpa | Elite | 1.000 | 0.905 | 1.105 |
| hscalc | Catch-up | 0.904 | 0.860 | 0.950 |
| hscalc | Elite | 1.049 | 0.997 | 1.104 |
| hsengl | Catch-up | 0.957 | 0.892 | 1.027 |
| hsengl | Elite | 0.942 | 0.896 | 0.991 |

# Beat The Blues Data

```
/* FinalBeatTheBlues.sas */
title 'Beat The Blues';

proc format;
     value ynfmt  0 = 'No'  1 = 'Yes';

data blues;
     infile '/folders/myfolders/BeatTheBlues.data.txt' firstobs=2;
     input id drug $ length $ treatment $ bdi_pre bdi_2m bdi_4m bdi_6m bdi_8m;
     label length = 'Length of Current Depression Episode';
     change2 = bdi_pre - bdi_2m; /* Change represents improvement. */
     change4 = bdi_pre - bdi_4m;
     change6 = bdi_pre - bdi_6m;
     change8 = bdi_pre - bdi_8m;


data blues2;
     set blues;
     Time = 2; Change = change2; output;
     Time = 4; Change = change4; output;
     Time = 6; Change = change6; output;
     Time = 8; Change = change8; output;
     keep id drug treatment time change;

/* proc print; */

proc mixed cl data=blues2;
     class drug treatment time;
     model change = drug|treatment|time;
     repeated / type=un subject=id;
     lsmeans drug treatment time;
```

## The Mixed Procedure

| Model Information | |
|---|---|
| **Data Set** | WORK.BLUES2 |
| **Dependent Variable** | Change |
| **Covariance Structure** | Unstructured |
| **Subject Effect** | id |
| **Estimation Method** | REML |
| **Residual Variance Method** | None |
| **Fixed Effects SE Method** | Model-Based |
| **Degrees of Freedom Method** | Between-Within |

| Class Level Information | | |
|---|---|---|
| **Class** | **Levels** | **Values** |
| **drug** | 2 | No Yes |
| **treatment** | 2 | BtheB TAU |
| **Time** | 4 | 2 4 6 8 |

| Dimensions | |
|---|---|
| **Covariance Parameters** | 10 |
| **Columns in X** | 45 |
| **Columns in Z** | 0 |
| **Subjects** | 100 |
| **Max Obs per Subject** | 4 |

| Number of Observations | |
|---|---|
| **Number of Observations Read** | 400 |
| **Number of Observations Used** | 280 |
| **Number of Observations Not Used** | 120 |

| Iteration History | | | |
|---|---|---|---|
| **Iteration** | **Evaluations** | **-2 Res Log Like** | **Criterion** |
| **0** | 1 | 1988.26231837 | |
| **1** | 2 | 1832.29955627 | 0.00400833 |
| **2** | 1 | 1829.11377367 | 0.00049929 |
| **3** | 1 | 1828.74707488 | 0.00001321 |
| **4** | 1 | 1828.73801079 | 0.00000001 |
| **5** | 1 | 1828.73800078 | 0.00000000 |

## The Mixed Procedure

Convergence criteria met.

| Covariance Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| Cov Parm | Subject | Estimate | Alpha | Lower | Upper |
| UN(1,1) | id | 84.5856 | 0.05 | 64.7064 | 115.32 |
| UN(2,1) | id | 62.8312 | 0.05 | 39.3372 | 86.3252 |
| UN(2,2) | id | 97.0287 | 0.05 | 72.2035 | 137.36 |
| UN(3,1) | id | 63.8506 | 0.05 | 40.2906 | 87.4107 |
| UN(3,2) | id | 70.5474 | 0.05 | 42.9462 | 98.1486 |
| UN(3,3) | id | 89.3646 | 0.05 | 64.6702 | 131.57 |
| UN(4,1) | id | 66.5065 | 0.05 | 41.3092 | 91.7039 |
| UN(4,2) | id | 72.2143 | 0.05 | 42.4840 | 101.94 |
| UN(4,3) | id | 74.7155 | 0.05 | 44.5305 | 104.90 |
| UN(4,4) | id | 101.73 | 0.05 | 72.7596 | 152.34 |

| Fit Statistics | |
|---|---|
| -2 Res Log Likelihood | 1828.7 |
| AIC (Smaller is Better) | 1848.7 |
| AICC (Smaller is Better) | 1849.6 |
| BIC (Smaller is Better) | 1874.8 |

| Null Model Likelihood Ratio Test | | |
|---|---|---|
| DF | Chi-Square | Pr > ChiSq |
| 9 | 159.52 | <.0001 |

| Type 3 Tests of Fixed Effects | | | | |
|---|---|---|---|---|
| Effect | Num DF | Den DF | F Value | Pr > F |
| drug | 1 | 93 | 4.82 | 0.0306 |
| treatment | 1 | 93 | 0.06 | 0.7999 |
| drug*treatment | 1 | 93 | 1.51 | 0.2230 |
| Time | 3 | 93 | 7.03 | 0.0003 |
| drug*Time | 3 | 93 | 0.66 | 0.5802 |
| treatment*Time | 3 | 93 | 0.84 | 0.4752 |
| drug*treatment*Time | 3 | 93 | 2.67 | 0.0521 |

## The Mixed Procedure

**Least Squares Means**

| Effect | drug | treatment | Time | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|---|---|
| **drug** | No | | | 6.6349 | 1.2626 | 93 | 5.25 | <.0001 |
| **drug** | Yes | | | 10.9960 | 1.5329 | 93 | 7.17 | <.0001 |
| **treatment** | | BtheB | | 9.0679 | 1.2916 | 93 | 7.02 | <.0001 |
| **treatment** | | TAU | | 8.5630 | 1.5085 | 93 | 5.68 | <.0001 |
| **Time** | | | 2 | 6.5811 | 1.0086 | 93 | 6.52 | <.0001 |
| **Time** | | | 4 | 8.0684 | 1.1619 | 93 | 6.94 | <.0001 |
| **Time** | | | 6 | 9.7358 | 1.1511 | 93 | 8.46 | <.0001 |
| **Time** | | | 8 | 10.8764 | 1.2654 | 93 | 8.60 | <.0001 |