Generalized Linear Mixed Models for Everything

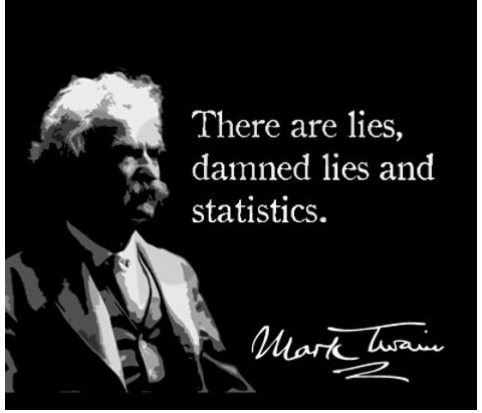
BERDC Special Topics Talk 2

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Dr. Mark Williamson

Biostatistics, Epidemiology, and Research Design Core

Introduction



Statistics

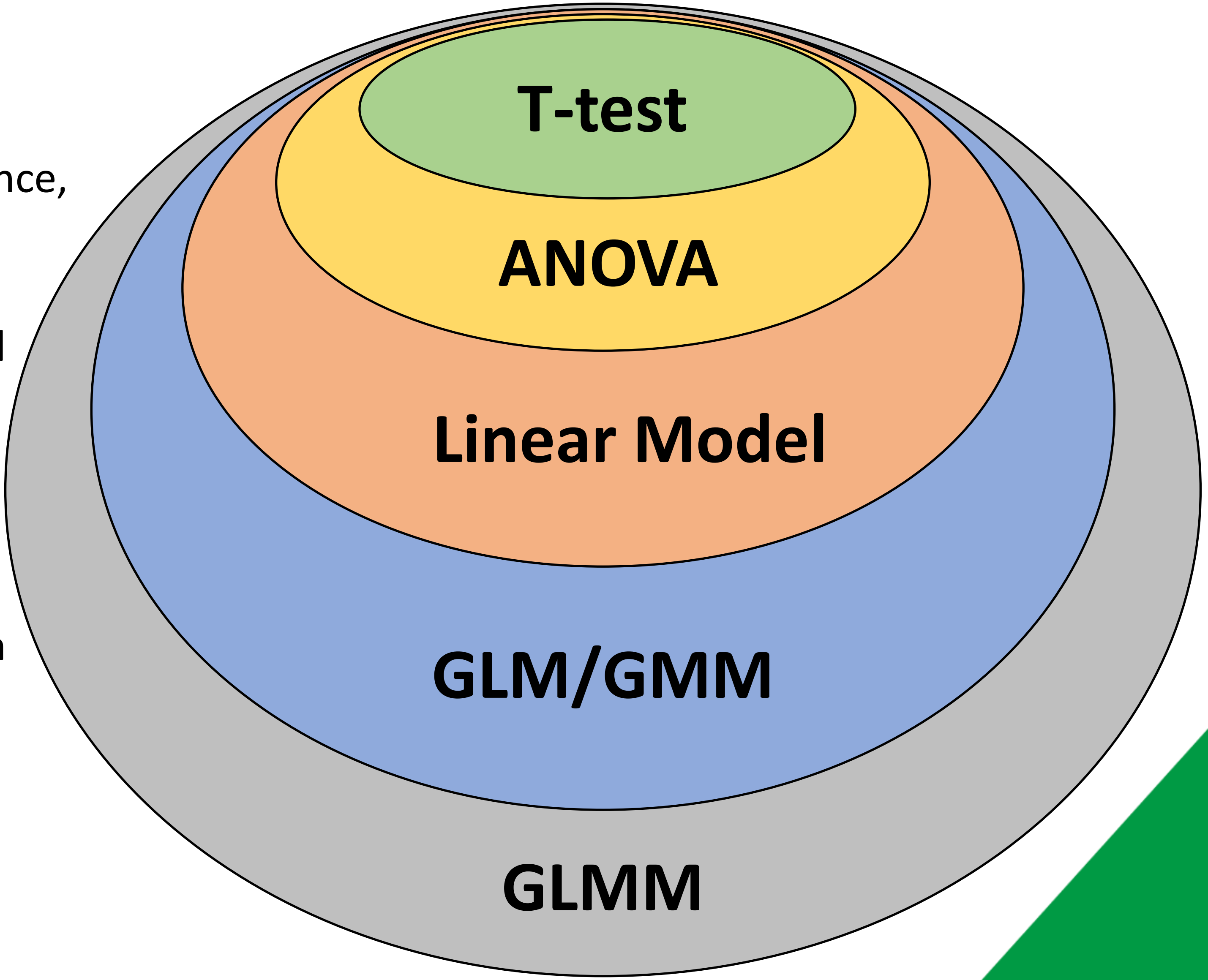
- Using mathematical methods for inference, estimation, and prediction of collected data
- Using mathematics to help discover and tell the truth about the world

Old way of approaching statistics

- By hand or limited computing
- If response variable not normally distributed, make it close enough or run non-parametric

New way of approaching statistics

- Given new understanding and computing power, can flexibly fit a vast range of data types using a unified framework



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND
UNIVERSITY OF NORTH DAKOTA

What are GLMMs?

- Generalized Linear Mixed Model
- Model that allows for non-normally distributed response variables (y) and predictor variables (x) as fixed and/or random effects
  - Non-normal: binary, binomial, beta, Poisson, negative binomial, exponential, log-normal, gamma, etc.
  - Fixed effect: Categorical variable in which all levels of interest are included
  - Random effect: Categorical variable in which levels included are subset of all levels

Egg_Counts	=	Temperature	+	Species	+	Sandbar
Poisson		*Num*		*Fixed*		*Random*

Disease_Status	=	Age_Class	+	Sex	+	Ethnicity	+	Hospital
Binary		*Fixed*		*Fixed*		*Fixed*		*Random*

Grade	=	Math_Score	+	Reading_Score	+	Classroom	+	School
Beta		*Num*		*Num*		*Random*		*Random*

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

PROC GLIMMIX (SAS)

Basic Syntax

```
PROC GLIMMIX data=Dataset;
        class Cat1 Cat2 Block;
        model Y=Num1 Cat1|Cat2
                /dist=Dist;
        random Block;
```

Options

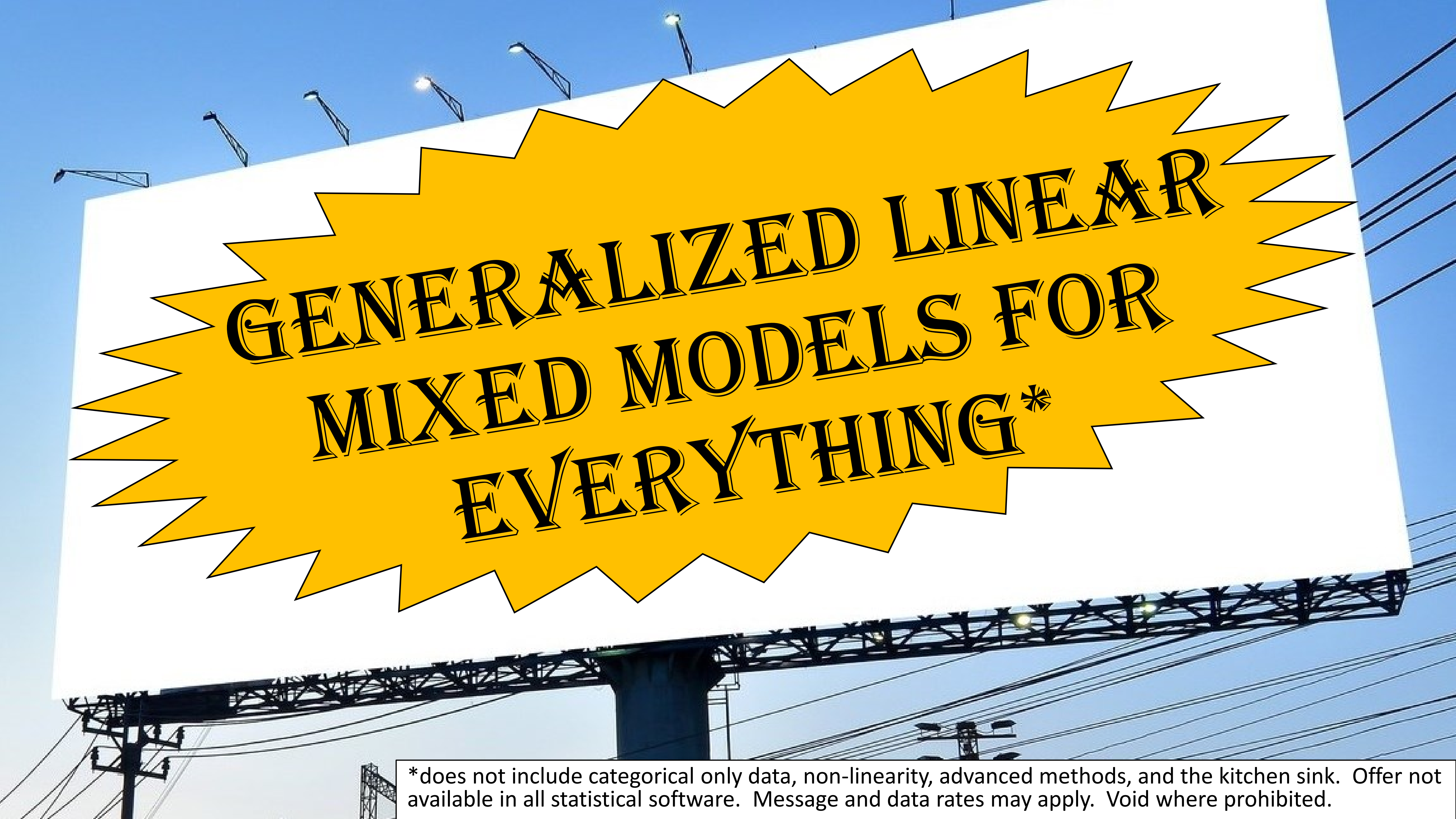
```
PROC GLIMMIX data=Dataset method=___ plots=___;
        by ___;
        where ____;
        model Y=Num1 Cat1|Cat2 / dist=Dist
                solution oddsratio link=___ offset=___;
        random intercept Cat1 /subject=Block;
        covtest ___;
        weight ___;
        lsmeans Cat1 / ilink cl;
        ods output LSMeans=lsm_dataset;
        output out=pred_dataset pred(ilink) lcl(ilink) ucl(ilink);
```

Resources

- https://support.sas.com/documentation/onlinedoc/stat/131/glimmix.pdf
- https://documentation.sas.com/?cdcId=statcdc&cdcVersion=14.2&docsetId=statug&docsetTarget=statug_glimmix_toc.htm&locale=en

Tests covered

- T-test
- ANOVA
- Linear regression
- Poisson regression
- Logistic regression
- Mixed model
- Generalized linear mixed model



*does not include categorical only data, non-linearity, advanced methods, and the kitchen sink. Offer not available in all statistical software. Message and data rates may apply. Void where prohibited.

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Example 1: T-tests

Dataset: sashelp.Class

- Includes Name, Sex and Height of 19 students

Question: Is there a difference in height between boys and girls?

```
PROC TTEST data=sashelp.Class;
        class Sex;
        var Height;
```

```
PROC GLIMMIX data=sashelp.Class;
        class Sex;
        model Height=Sex /solution dist=normal;
        lsmeans Sex /cl;
        ods output LSmeans=Class_lsm;
PROC SGPLOT data=Class_lsm;
        vbarparm category=Sex
        response=Estimate/limitupper=Upper
        limitlower=Lower;
```

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

T-test Results

The TTEST Procedure

Variable: Height

Sex	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
F		9	60.5889	5.0183	1.6728	51.3000	66.5000
M		10	63.9100	4.9379	1.5615	57.3000	72.0000
Diff (1-2)	Pooled		-3.3211	4.9759	2.2863		
Diff (1-2)	Satterthwaite		-3.3211		2.2883		

Sex	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
F		60.5889	56.7315	64.4463	5.0183	3.3897	9.6140
M		63.9100	60.3776	67.4424	4.9379	3.3965	9.0147
Diff (1-2)	Pooled	-3.3211	-8.1447	1.5025	4.9759	3.7339	7.4596
Diff (1-2)	Satterthwaite	-3.3211	-8.1551	1.5129			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	17	-1.45	0.1645
Satterthwaite	Unequal	16.727	-1.45	0.1652

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	8	9	1.03	0.9527

The GLIMMIX Procedure

Model Information

Data Set	SASHELP.CLASS
Response Variable	Height
Response Distribution	Gaussian
Link Function	Identity
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Residual

Class Level Information

Class	Levels	Values
Sex	2	F M

Number of Observations Read	19
Number of Observations Used	19

Dimensions

Covariance Parameters	1
Columns in X	3
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	19

Optimization Information

Optimization Technique	None
Parameters	3
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Not Profiled

Fit Statistics

-2 Res Log Likelihood	107.30
AIC (smaller is better)	113.30
AICC (smaller is better)	115.15
BIC (smaller is better)	115.80
CAIC (smaller is better)	118.80

Fit Statistics

HQIC (smaller is better)	113.55
Pearson Chi-Square	420.92
Pearson Chi-Square / DF	24.76

Parameter Estimates

Effect	Sex	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		63.9100	1.5735	17	40.62	<.0001
Sex	F	-3.3211	2.2863	17	-1.45	0.1645
Sex	M	0	.	.	.	.
Scale		24.7599	8.4926	.	.	.

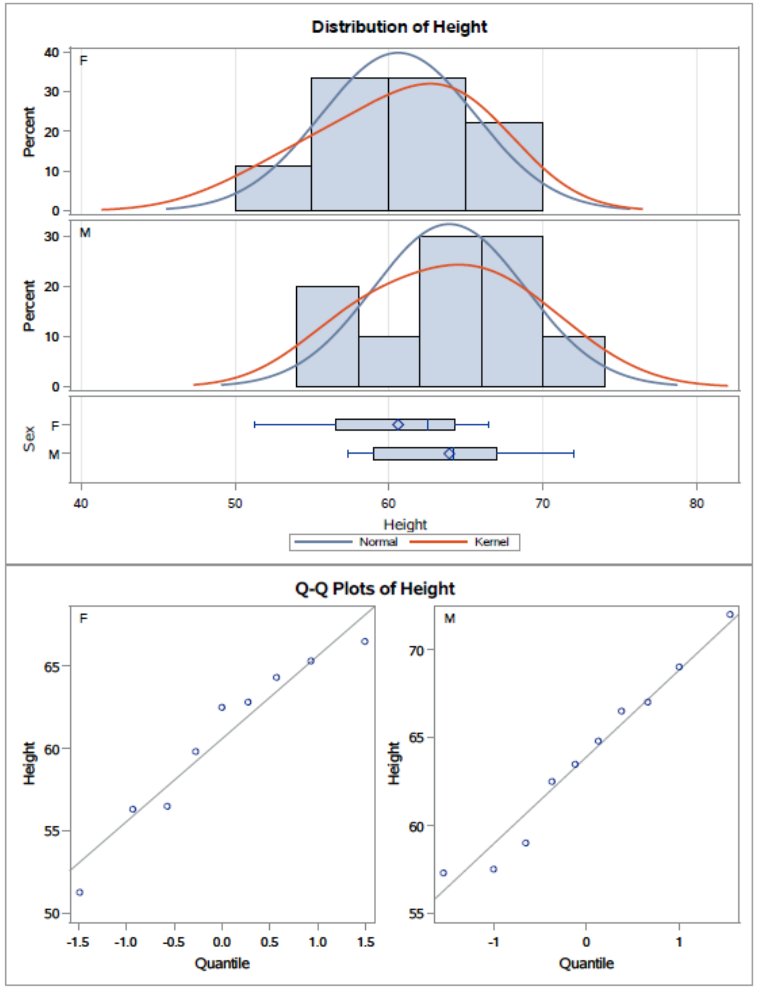
Type III Tests of Fixed Effects

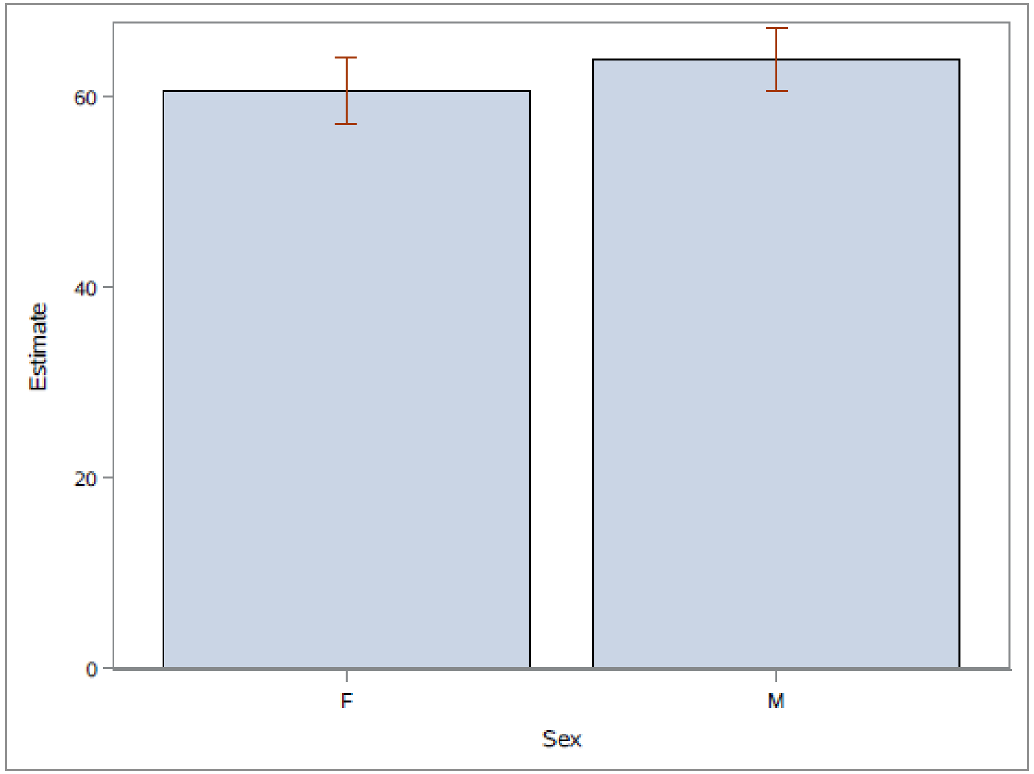
Effect	Num DF	Den DF	F Value	Pr > F
Sex	1	17	2.11	0.1645

Sex Least Squares Means

Sex	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
F	60.5889	1.6586	17	36.53	<.0001	0.05	57.0895	64.0883
M	63.9100	1.5735	17	40.62	<.0001	0.05	60.5901	67.2299

T-test Graphs





Example 2: ANOVA

Dataset: sashelp.bweight

- Includes infant birth weight (Weight) and mother's education level (MomEdLevel)

Question: Is there a difference in birthweight across education level?

```
PROC ANOVA data=sashelp.bweight;
        class MomEdLevel;
        model Weight=MomEdLevel;
        means MomEdLevel;
```

```
PROC GLIMMIX data=sashelp.bweight;
        class MomEdLevel;
        model Weight=MomEdLevel/solution dist=normal;
        lsmeans MomEdLevel / cl;
        ods output LSMeans=Bweight_lsm;
PROC SGPLOT data=Bweight_lsm;
        vbarparm category=MomEdLevel
        response=Estimate/
        limitupper=Upper limitlower=Lower;
```

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

ANOVA Results

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
MomEdLevel	4	0 1 2 3

Number of Observations Read	50000
Number of Observations Used	50000

Level of MomEdLevel	N	Weight Mean	Weight Std Dev
0	17449	3336.90412	579.656328
1	12129	3394.24330	564.161122
2	12449	3466.65772	531.045491
3	7973	3259.37451	567.353714

Dependent Variable: Weight Infant Birth Weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	240093609	80031203	253.26	<.0001
Error	49996	15799187158	316009		
Corrected Total	49999	16039280767			

R-Square	Coeff Var	Root MSE	Weight Mean
0.014969	16.67717	562.1468	3370.757

Source	DF	Anova SS	Mean Square	F Value	Pr > F
MomEdLevel	3	240093608.9	80031203.0	253.26	<.0001

The GLIMMIX Procedure

Model Information	
Data Set	SASHELP.BWEIGHT
Response Variable	Weight
Response Distribution	Gaussian
Link Function	Identity
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Residual

Class Level Information		
Class	Levels	Values
MomEdLevel	4	0 1 2 3

Fit Statistics	
-2 Res Log Likelihood	775045.7
AIC (smaller is better)	775055.7
AICC (smaller is better)	775055.7
BIC (smaller is better)	775099.8
CAIC (smaller is better)	775104.8
HQIC (smaller is better)	775069.5
Pearson Chi-Square	1.58E10
Pearson Chi-Square / DF	316009.0

Parameter Estimates

Effect	Mother's Education Level	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		3259.37	6.2956	49996	517.72	<.0001
MomEdLevel	0	77.5296	7.5990	49996	10.20	<.0001
MomEdLevel	1	134.87	8.1049	49996	16.64	<.0001
MomEdLevel	2	207.28	8.0634	49996	25.71	<.0001
MomEdLevel	3	0	.	.	.	.
Scale		316009	1998.70	.	.	.

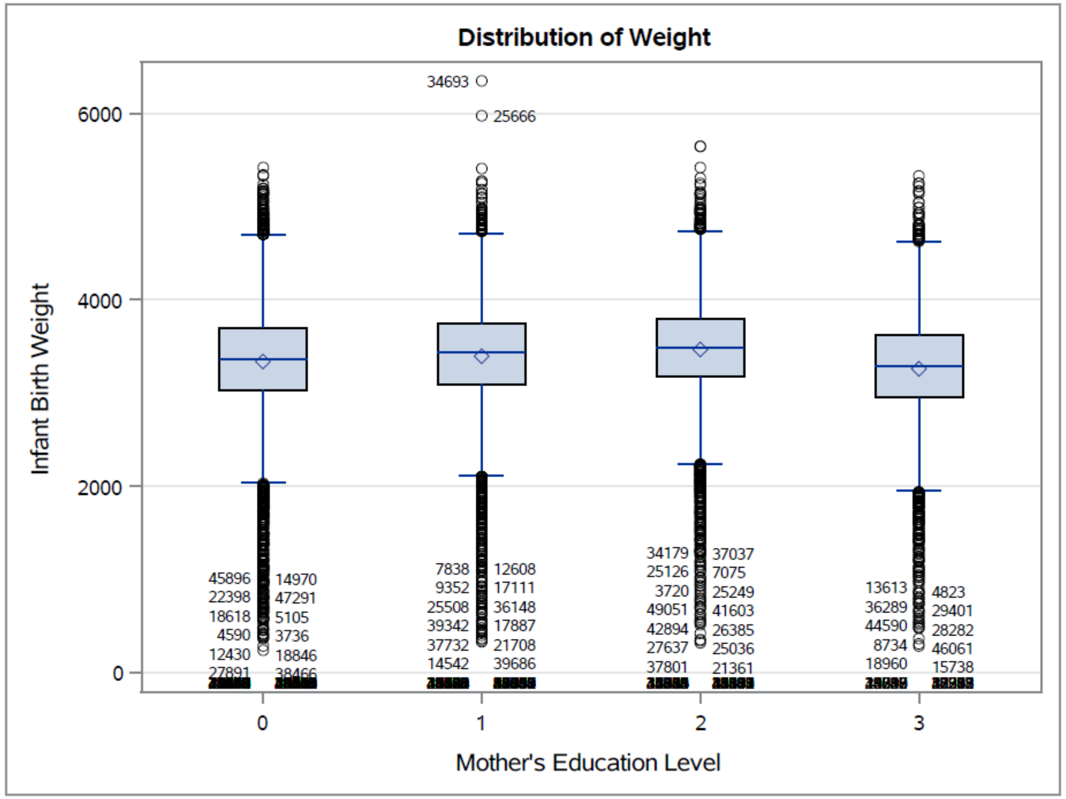
Type III Tests of Fixed Effects

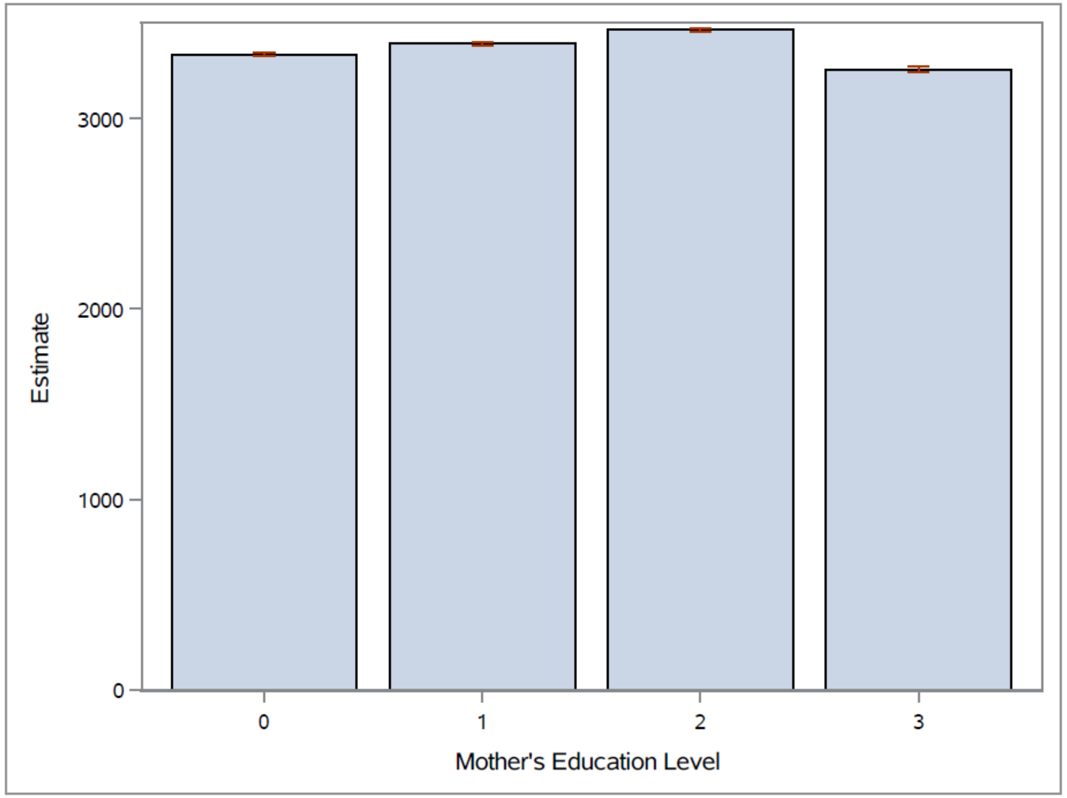
Effect	Num DF	Den DF	F Value	Pr > F
MomEdLevel	3	49996	253.26	<.0001

MomEdLevel Least Squares Means

Mother's Education Level	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
0	3336.90	4.2556	49996	784.11	<.0001	0.05	3328.56	3345.25
1	3394.24	5.1043	49996	664.98	<.0001	0.05	3384.24	3404.25
2	3466.66	5.0383	49996	688.06	<.0001	0.05	3456.78	3476.53
3	3259.37	6.2956	49996	517.72	<.0001	0.05	3247.04	3271.71

ANOVA Graphs





Example 3: Linear Regression

Dataset: sashelp.bmimen

- Includes BMI and Age for men

Question: Is there a relationship between BMI and age in men?

```
PROC REG data=sashelp.bmimen;
	model BMI=age;
```

```
PROC GLIMMIX data=sashelp.bmimen;
	model BMI=age/solution dist=normal;
	output out=Bmimen_pred pred lcl ucl;
PROC SGPLOT data=Bmimen_pred;
	band x=age lower=lcl upper=ucl;
	scatter x=age y=BMI;
	series x=age y=Pred;
```

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

Linear Regression Results

The REG Procedure
Model: MODEL1
Dependent Variable: BMI

Number of Observations Read	3264
Number of Observations Used	3264

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	58653	58653	2258.27	<.0001
Error	3262	84722	25.97229		
Corrected Total	3263	143374			

Root MSE	5.09630	R-Square	0.4091
Dependent Mean	22.11201	Adj R-Sq	0.4089
Coeff Var	23.04766		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	17.90619	0.12566	142.50	<.0001
Age	1	0.16242	0.00342	47.52	<.0001

The GLIMMIX Procedure

Model Information

Data Set	SASHELP.BMIMEN
Response Variable	BMI
Response Distribution	Gaussian
Link Function	Identity
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Residual

Fit Statistics

-2 Res Log Likelihood	19904.29
AIC (smaller is better)	19910.29
AICC (smaller is better)	19910.30
BIC (smaller is better)	19928.56
CAIC (smaller is better)	19931.56
HQIC (smaller is better)	19916.84
Pearson Chi-Square	84721.61
Pearson Chi-Square / DF	25.97

Parameter Estimates

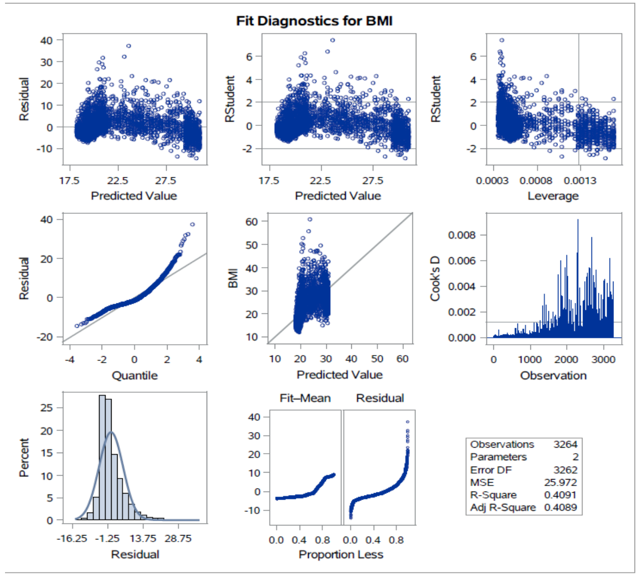
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	17.9062	0.1257	3262	142.50	<.0001
Age	0.1624	0.003418	3262	47.52	<.0001
Scale	25.9723	0.6431	.	.	.

Type III Tests of Fixed Effects

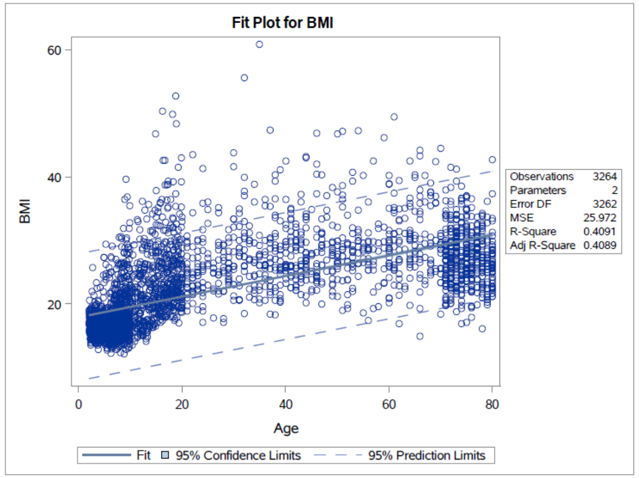
Effect	Num DF	Den DF	F Value	Pr > F
Age	1	3262	2258.27	<.0001

Linear Regression Graphs

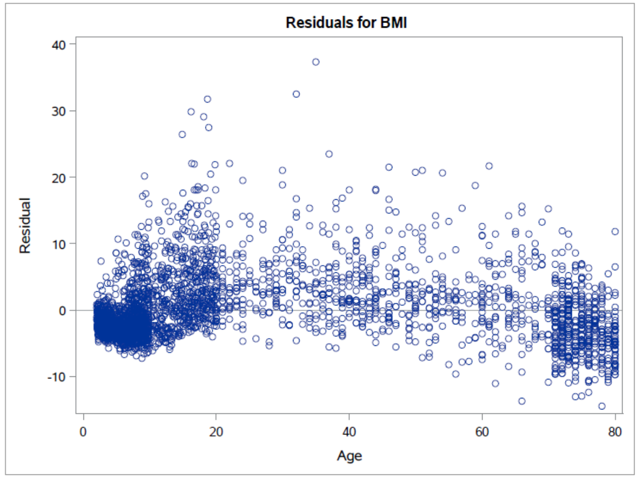
The REG Procedure
Model: MODEL1
Dependent Variable: BMI

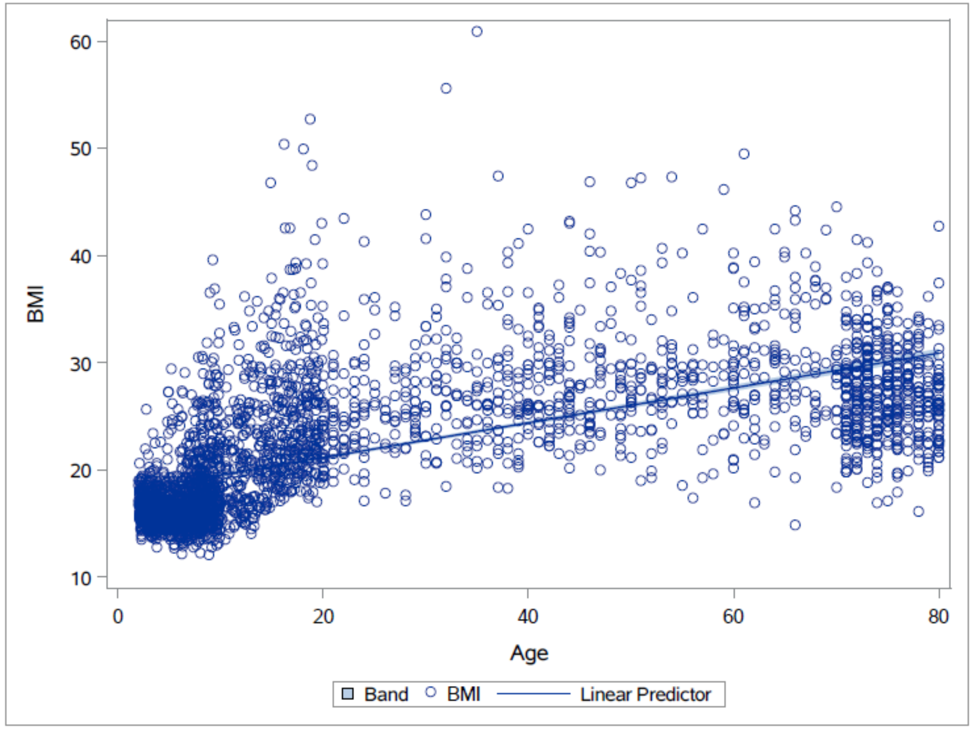


The REG Procedure
Model: MODEL1
Dependent Variable: BMI



The REG Procedure
Model: MODEL1
Dependent Variable: BMI





Example 4: Logistic Regression

Dataset: sashelp.birthwgt

- Includes infant birth weight (LowBirthWght: Yes/No), marriage status (Married), age group (AgeGroup) and race (Race: White, Asian, Black, Hispanic, Native)

Question: Does marriage status, age group, and race help predict the probability of having an infant of low birth weight?

```
PROC LOGISTIC data=sashelp.birthwgt;
        class Married AgeGroup Race;
        model LowBirthWgt(Event='Yes’)=
                Married AgeGroup Race;
```

```
PROC GLIMMIX data=sashelp.birthwgt;
        class Married AgeGroup Race;
        model LowBirthWgt(Event='Yes’)=
                Married AgeGroup Race /
                solution dist=binary oddsratio;
```

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

Logistic Regression Results

The LOGISTIC Procedure

Model Information		
Data Set	SASHELP.BIRTHWGT	Mediation Effect of Low Birth Weight on Infant Mortality
Response Variable	LowBirthWgt	
Number of Response Levels	2	
Model	binary logit	
Optimization Technique	Fisher's scoring	

Number of Observations Read	100000
Number of Observations Used	100000

Response Profile		
Ordered Value	LowBirthWgt	Total Frequency
1	No	91859
2	Yes	8141

Probability modeled is LowBirthWgt='Yes'.

Class Level Information					
Class	Value	Design Variables			
Married	No	1			
	Yes	-1			
AgeGroup	1	1	0		
	2	0	1		
	3	-1	-1		
Race	Asian	1	0	0	0
	Black	0	1	0	0
	Hispanic	0	0	1	0
	Native	0	0	0	1
	White	-1	-1	-1	-1

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

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	56441.929	55527.740
SC	56451.442	55603.843
-2 Log L	56439.929	55511.740

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	928.1894	7	<.0001
Score	1042.7339	7	<.0001
Wald	1001.9075	7	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Married	1	134.9503	<.0001
AgeGroup	2	79.9405	<.0001
Race	4	537.3870	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.2961	0.0294	6110.4752	<.0001
Married	No	1	-0.1573	0.0135	134.9503	<.0001
AgeGroup	1	1	-0.0423	0.0263	2.5856	0.1078
AgeGroup	2	1	-0.1232	0.0174	49.8798	<.0001
Race	Asian	1	-0.0216	0.0494	0.1907	0.6623
Race	Black	1	0.5165	0.0340	230.8965	<.0001
Race	Hispanic	1	-0.2293	0.0345	44.0655	<.0001
Race	Native	1	-0.1456	0.0986	2.1830	0.1395

The GLIMMIX Procedure

Model Information	
Data Set	SASHELP.BIRTHWGT
Response Variable	LowBirthWgt
Response Distribution	Binary
Link Function	Logit
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Maximum Likelihood
Degrees of Freedom Method	Residual

Class Level Information		
Class	Levels	Values
Married	2	No Yes
AgeGroup	3	1 2 3
Race	5	Asian Black Hispanic Native White

Number of Observations Read	100000
Number of Observations Used	100000

Response Profile		
Ordered Value	LowBirthWgt	Total Frequency
1	No	91859
2	Yes	8141

The GLIMMIX procedure is modeling the probability that LowBirthWgt='Yes'.

Fit Statistics	
-2 Log Likelihood	55511.74
AIC (smaller is better)	55527.74
AICC (smaller is better)	55527.74
BIC (smaller is better)	55603.84
CAIC (smaller is better)	55611.84
HQIC (smaller is better)	55550.84
Pearson Chi-Square	99989.11
Pearson Chi-Square / DF	1.00

Parameter Estimates								
Effect	Married	Race	AgeGroup	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept				-2.0933	0.03895	99992	-53.74	<.0001
Married	No			-0.3146	0.02709	99992	-11.62	<.0001
Married	Yes			0	.	.	.	.
AgeGroup			1	-0.2078	0.04722	99992	-4.40	<.0001
AgeGroup			2	-0.2887	0.03260	99992	-8.86	<.0001
AgeGroup			3	0	.	.	.	.
Race		Asian		0.09846	0.05443	99992	1.81	0.0705
Race		Black		0.6365	0.03119	99992	20.41	<.0001
Race		Hispanic		-0.1093	0.03176	99992	-3.44	0.0006
Race		Native		-0.02562	0.1234	99992	-0.21	0.8356
Race		White		0	.	.	.	.

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Married	1	99992	134.95	<.0001
AgeGroup	2	99992	39.97	<.0001
Race	4	99992	134.32	<.0001

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

Logistic Regression Tables

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Married No vs Yes	0.730	0.692	0.770
AgeGroup 1 vs 3	0.812	0.741	0.891
AgeGroup 2 vs 3	0.749	0.703	0.799
Race Asian vs White	1.103	0.992	1.228
Race Black vs White	1.890	1.778	2.009
Race Hispanic vs White	0.897	0.842	0.954
Race Native vs White	0.975	0.765	1.241

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	52.1	Somers' D	0.180
Percent Discordant	34.1	Gamma	0.209
Percent Tied	13.7	Tau-a	0.027
Pairs	747824119	c	0.590

Odds Ratio Estimates									
Married	AgeGroup	Race	_Married	_AgeGroup	_Race	Estimate	DF	95% Confidence Limits	
No			Yes			0.730	99992	0.692	0.770
	1			3		0.812	99992	0.741	0.891
	2			3		0.749	99992	0.703	0.799
		Asian			White	1.103	99992	0.992	1.228
		Black			White	1.890	99992	1.778	2.009
		Hispanic			White	0.897	99992	0.842	0.954
		Native			White	0.975	99992	0.765	1.241

Probability modeled is LowBirthWgt='Yes'.

Example 5: Poisson Regression

Dataset: poisson.sas7bdat

- https://stats.idre.ucla.edu/sas/output/poisson-regression/
- Includes days absent during the school year (daysabs), math standardized tests score (mathnce), language standardized tests score (mangnce), and gender (female)

Question: Do math scores, language scores, and gender help predict the number of days absent from school?

```
PROC GENMOD data=absent;
        model daysabs=mathnce
        langnce female /
        link=log dist=Poisson;
```

```
PROC GLIMMIX data=absent;
        model daysabs=mathnce
        langnce female /
        solution dist=Poisson;
```

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Poisson Regression Results

The GENMOD Procedure

Model Information		
Data Set	WORK.ABSENT	
Distribution	Poisson	
Link Function	Log	
Dependent Variable	DAYSABS	number days absent

Number of Observations Read	316
Number of Observations Used	316

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	312	2234.5462	7.1620
Scaled Deviance	312	2234.5462	7.1620
Pearson Chi-Square	312	2774.4139	8.8924
Scaled Pearson X2	312	2774.4139	8.8924
Log Likelihood		1482.2670	
Full Log Likelihood		-1547.9709	
AIC (smaller is better)		3103.9419	
AICC (smaller is better)		3104.0705	
BIC (smaller is better)		3118.9649	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	2.2867	0.0700	2.1496	2.4239	1068.59	<.0001
MATHNCE	1	-0.0035	0.0018	-0.0071	0.0000	3.74	0.0531
LANGNCE	1	-0.0122	0.0018	-0.0157	-0.0086	43.86	<.0001
female	1	0.4009	0.0484	0.3060	0.4958	68.58	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.

The GLIMMIX Procedure

Model Information	
Data Set	WORK.ABSENT
Response Variable	DAYSABS
Response Distribution	Poisson
Link Function	Log
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Maximum Likelihood
Degrees of Freedom Method	Residual

Fit Statistics	
-2 Log Likelihood	3095.94
AIC (smaller is better)	3103.94
AICC (smaller is better)	3104.07
BIC (smaller is better)	3118.96
CAIC (smaller is better)	3122.96
HQIC (smaller is better)	3109.94
Pearson Chi-Square	2774.41
Pearson Chi-Square / DF	8.89

Parameter Estimates

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	2.2867	0.06995	312	32.69	<.0001
MATHNCE	-0.00352	0.001821	312	-1.93	0.0540
LANGNCE	-0.01215	0.001835	312	-6.62	<.0001
female	0.4009	0.04841	312	8.28	<.0001

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
MATHNCE	1	312	3.74	0.0540
LANGNCE	1	312	43.86	<.0001
female	1	312	68.58	<.0001

Example 6: Linear Mixed Models

Dataset: Split-Plot

- https://documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.4&docsetId=statug&docsetTarget=statug_mixed_examples01.htm&locale=en
- Includes block (Block), whole-plot factor (A), subplot factor (B), and yield (Y)

Question: Is yield different across a whole-plot factor and subplot factor, while controlling for blocking?

```
PROC MIXED data=sp;
        class A B Block;
        model Y = A B A*B;
        random Block A*Block/solution;
```

```
PROC GLIMMIX data=sp;
        class A B Block;
        model Y = A B A*B;
        random Block A*Block/solution;
```

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

Linear Mixed Models Results

The Mixed Procedure

Model Information	
Data Set	WORK.SP
Dependent Variable	Y
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
A	3	1 2 3
B	2	1 2
Block	4	1 2 3 4

Dimensions	
Covariance Parameters	3
Columns in X	12
Columns in Z	16
Subjects	1
Max Obs per Subject	24

Number of Observations	
Number of Observations Read	24
Number of Observations Used	24
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	139.81461222	
1	1	119.76184570	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate
Block	62.3958
A*Block	15.3819
Residual	9.3611

Fit Statistics	
-2 Res Log Likelihood	119.8
AIC (Smaller is Better)	125.8
AICC (Smaller is Better)	127.5
BIC (Smaller is Better)	123.9

Solution for Random Effects

Effect	A	Block	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Block		1	10.7631	4.4865	9	2.40	0.0400
Block		2	-0.5269	4.4865	9	-0.12	0.9091
Block		3	-5.6450	4.4865	9	-1.26	0.2400
Block		4	-4.5912	4.4865	9	-1.02	0.3329
A*Block	1	1	3.7276	3.0331	9	1.23	0.2502
A*Block	1	2	-3.7171	3.0331	9	-1.23	0.2515
A*Block	1	3	0.5903	3.0331	9	0.19	0.8500
A*Block	1	4	-0.6009	3.0331	9	-0.20	0.8474
A*Block	2	1	-1.4476	3.0331	9	-0.48	0.6446
A*Block	2	2	-1.2253	3.0331	9	-0.40	0.6957
A*Block	2	3	0.3987	3.0331	9	0.13	0.8983
A*Block	2	4	2.2742	3.0331	9	0.75	0.4725
A*Block	3	1	0.3733	3.0331	9	0.12	0.9047
A*Block	3	2	4.8125	3.0331	9	1.59	0.1471
A*Block	3	3	-2.3806	3.0331	9	-0.78	0.4527
A*Block	3	4	-2.8052	3.0331	9	-0.92	0.3792

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
A	2	6	4.07	0.0764
B	1	9	19.39	0.0017
A*B	2	9	4.02	0.0566

The GLIMMIX Procedure

Model Information	
Data Set	WORK.SP
Response Variable	Y
Response Distribution	Gaussian
Link Function	Identity
Variance Function	Default
Variance Matrix	Not blocked
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
A	3	1 2 3
B	2	1 2
Block	4	1 2 3 4

Fit Statistics	
-2 Res Log Likelihood	119.76
AIC (smaller is better)	125.76
AICC (smaller is better)	127.48
BIC (smaller is better)	123.92
CAIC (smaller is better)	126.92
HQIC (smaller is better)	121.72
Generalized Chi-Square	168.50
Gener. Chi-Square / DF	9.36

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
Block	62.3958	56.5383
A*Block	15.3819	11.7914
Residual	9.3611	4.4129

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
A	2	6	4.07	0.0764
B	1	9	19.39	0.0017
A*B	2	9	4.02	0.0566

Solution for Random Effects

Effect	A	Block	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Block		1	10.7631	4.4865	9	2.40	0.0400
Block		2	-0.5269	4.4865	9	-0.12	0.9091
Block		3	-5.6450	4.4865	9	-1.26	0.2400
Block		4	-4.5912	4.4865	9	-1.02	0.3329
A*Block	1	1	3.7276	3.0331	9	1.23	0.2502
A*Block	1	2	-3.7171	3.0331	9	-1.23	0.2515
A*Block	1	3	0.5903	3.0331	9	0.19	0.8500
A*Block	1	4	-0.6009	3.0331	9	-0.20	0.8474
A*Block	2	1	-1.4476	3.0331	9	-0.48	0.6446
A*Block	2	2	-1.2253	3.0331	9	-0.40	0.6957
A*Block	2	3	0.3987	3.0331	9	0.13	0.8983
A*Block	2	4	2.2742	3.0331	9	0.75	0.4725
A*Block	3	1	0.3733	3.0331	9	0.12	0.9047
A*Block	3	2	4.8125	3.0331	9	1.59	0.1471
A*Block	3	3	-2.3806	3.0331	9	-0.78	0.4527
A*Block	3	4	-2.8052	3.0331	9	-0.92	0.3792

Example 7: Actual GLMMs

Dataset: poisson.sas7bdat

- https://support.sas.com/resources/papers/proceedings/proceedings/sugi30/196-30.pdf
- Includes hospital center (center), treatment group (group), total number of individuals (n) and number of individuals with side effect (SideEffect)

Question: Is the proportion of those with side effects different between treatment groups while controlling for hospital center?

```
PROC GLIMMIX data=multicenter;
        class center group;
        model SideEffect/n = group / solution;
        random center;
        lsmeans group / ilink cl;
        ods output LSMeans=lsm1;
```

```
PROC SGPLOT data=lsm1;
        vbarparm category=group
        response=Mu /
        limitlower=LowerMu
        limitupper=UpperMu;
```

Actual GLMMs Results

The GLIMMIX Procedure

Model Information	
Data Set	WORK.MULTICENTER
Response Variable (Events)	SideEffect
Response Variable (Trials)	n
Response Distribution	Binomial
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	center
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
center	15	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
group	2	A B

Number of Observations Read	30
Number of Observations Used	30
Number of Events	155
Number of Trials	503

Dimensions	
G-side Cov. Parameters	1
Columns in X	3
Columns in Z per Subject	1
Subjects (Blocks in V)	15
Max Obs per Subject	2

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	1
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Profiled
Starting From	Data

Iteration History

Iteration	Restarts	Subiterations	Objective Function	Change	Max Gradient
0	0	5	79.688580269	0.11807224	7.851E-7
1	0	3	81.294622554	0.02558021	8.209E-7
2	0	2	81.438701534	0.00166079	4.061E-8
3	0	1	81.444083567	0.00006263	2.273E-8
4	0	1	81.444265216	0.00000421	0.000025
5	0	1	81.444277364	0.00000383	0.000023
6	0	1	81.444266322	0.00000348	0.000021
7	0	1	81.44427636	0.00000316	0.000019
8	0	1	81.444267235	0.00000287	0.000017
9	0	1	81.44427553	0.00000261	0.000016
10	0	1	81.44426799	0.00000237	0.000014
11	0	1	81.444274844	0.00000216	0.000013
12	0	1	81.444268614	0.00000196	0.000012
13	0	1	81.444274277	0.00000178	0.000011
14	0	1	81.444269129	0.00000162	9.772E-6
15	0	0	81.444273808	0.00000000	9.102E-6

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	81.44
Generalized Chi-Square	30.69
Gener. Chi-Square / DF	1.10

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Intercept	center	0.6176	0.3181

Solutions for Fixed Effects

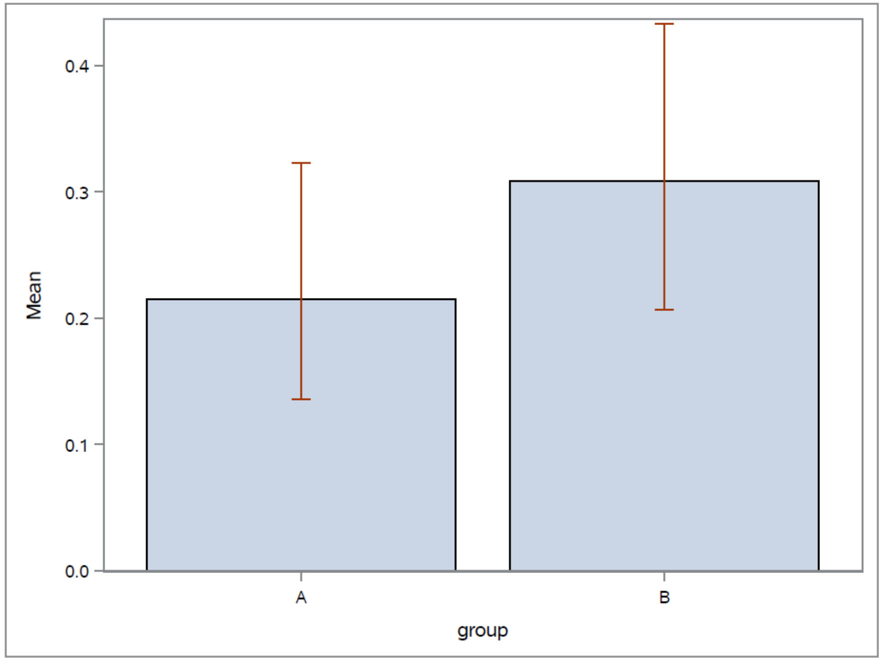
Effect	group	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		-0.8071	0.2514	14	-3.21	0.0063
group	A	-0.4896	0.2034	14	-2.41	0.0305
group	B	0	.	.	.	.

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
group	1	14	5.79	0.0305

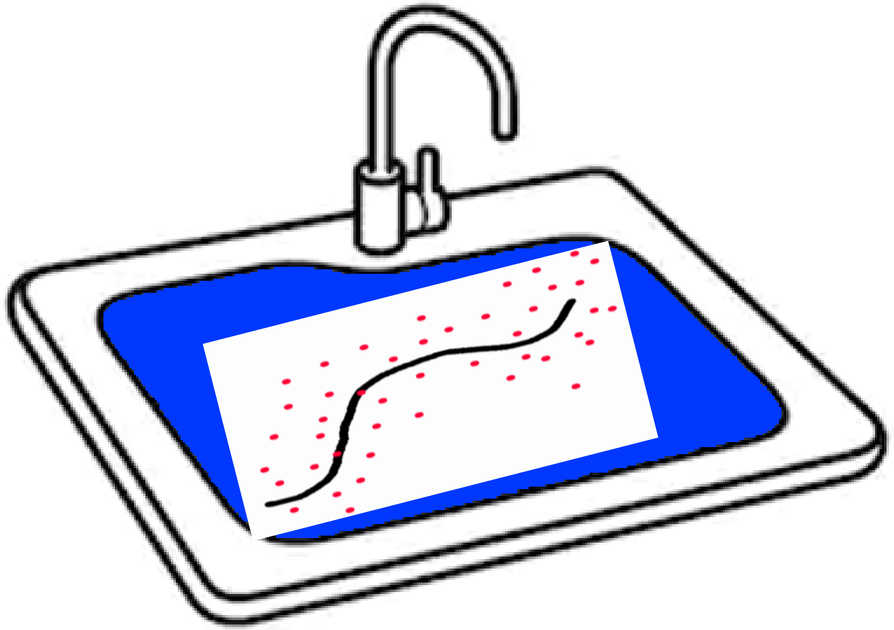
group Least Squares Means

group	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Mean	Standard Error Mean	Lower Mean	Upper Mean
A	-1.2966	0.2601	14	-4.99	0.0002	0.05	-1.8544	-0.7388	0.2147	0.04385	0.1354	0.3233
B	-0.8071	0.2514	14	-3.21	0.0063	0.05	-1.3462	-0.2679	0.3085	0.05363	0.2065	0.4334



The Kitchen Sink: Things you can't do

- Categorical only data (PROC FREQ)
- Non-linear models (PROC NLIN)
- Hierarchical models with non-normal random effects (PROC MCMC)
- Other specialized advanced models such as Structural Equation Modeling, Principal Components Analysis, etc.



DaCCoTA
DAKOTA CANCER COLLABORATIVE
ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Acknowledgements

- The DaCCoTA is supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729.
- For the labs that use the Biostatistics, Epidemiology, and Research Design Core in any way, including this Module, please acknowledge us for publications. ***"Research reported in this publication was supported by DaCCoTA (the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729)".***

DaCCoTA
DAKOTA CANCER COLLABORATIVE
ON TRANSLATIONAL ACTIVITY