ANOVA

Assumptions

$$Y_{ij} = \mu + T_i + \varepsilon_{(i)j}$$

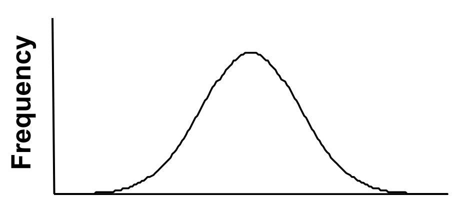
1) **Effects** are additive

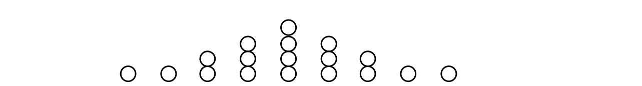
Errors are:

2) Normally distributed
3) Homogenous
4) Independent

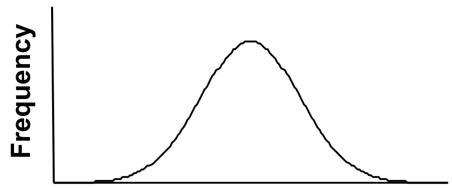
The Normal Distribution

$$f = \frac{N}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$





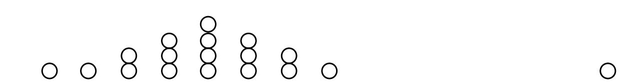
The Normal Distribution



Properties:
- Symmetric about mean
- Can be described by two variables; mean, standard deviation (s)
- Mean = median = mode
- Most commonly used; even for non-normal data

The Normal Distribution
Departures

Outliers

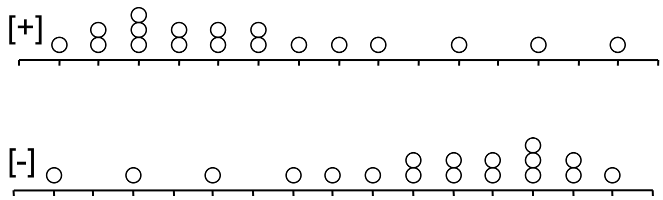


Identifying outliers
- 3*SD
- 1.5*IQR

The Normal Distribution

Departures

Skewness

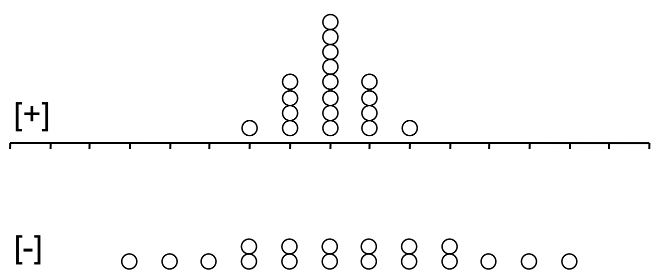


$$\gamma_1 = \frac{\sum (X_i - \mu)^3}{N\sigma^3}$$

The Normal Distribution

Departures

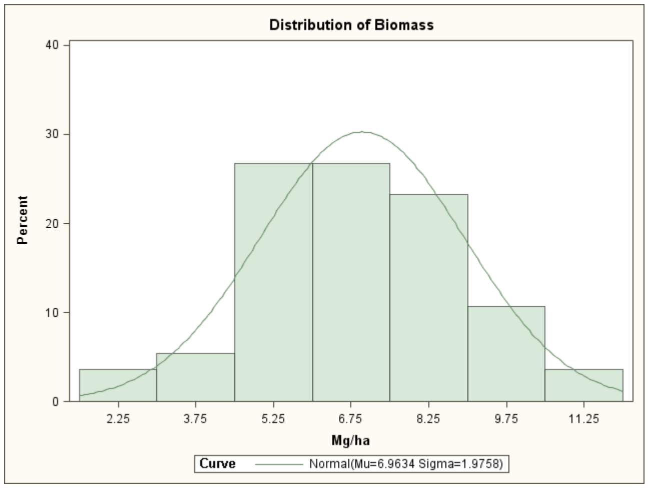
Kurtosis



$$\gamma_2 = \frac{\sum (X_i - \mu)^4}{N\sigma^4} - 3$$

Switchgrass Example

Biomass (t/ha)



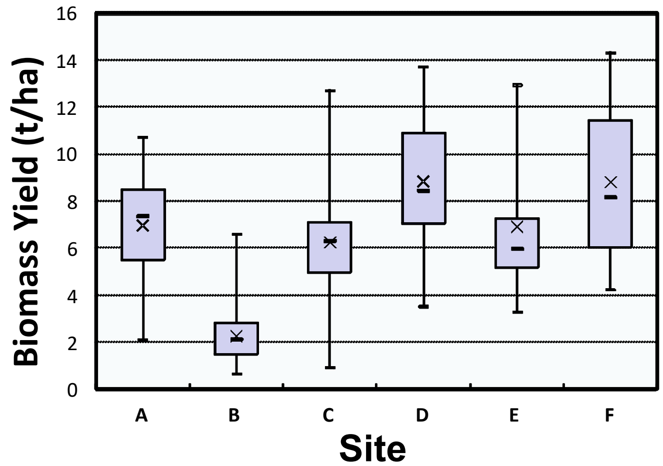
Switchgrass Example

Biomass (t/ha)

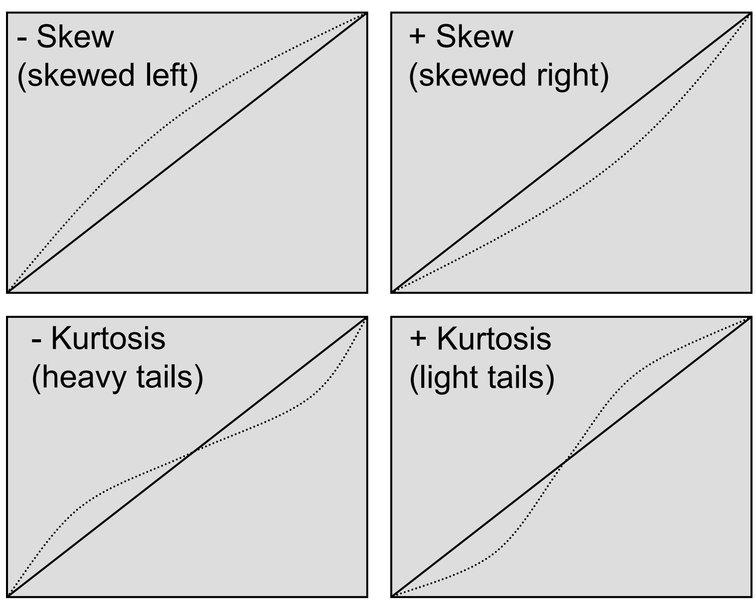
(t/ha)	
Mean	6.963431
Standard Error	0.26403
Median	7.34759
Mode	8.697148
Standard Deviation	1.97582
Sample Variance	3.903864
Kurtosis	-0.26821
Skewness	-0.32829
Range	8.597181
Minimum	2.099312
Maximum	10.69649
Sum	389.9521
Count	56

Assessing Normality

Box Plots

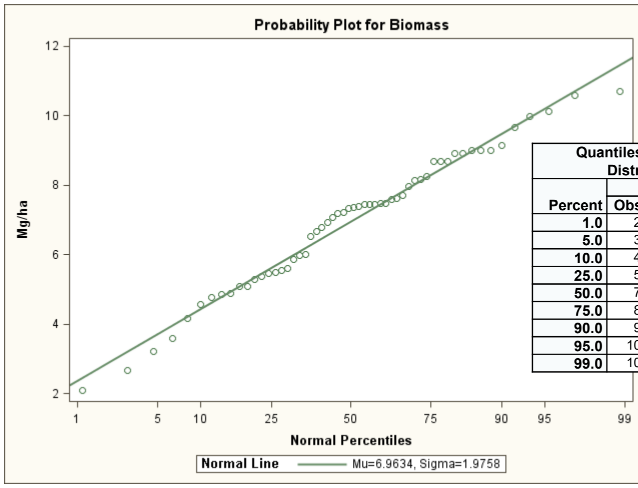


Normal Probability Plots



Assessing Normality

Normal Probability Plots



Quantiles for Normal Distribution		
	Quantile	
Percent	Observed	Estimated
1.0	2.09931	2.36699
5.0	3.21894	3.71350
10.0	4.57850	4.43132
25.0	5.47820	5.63076
50.0	7.34759	6.96343
75.0	8.47722	8.29610
90.0	9.15700	9.49555
95.0	10.11668	10.21337
99.0	10.69649	11.55987

Testing for Normality

- Shapiro-Wilk *W* test
- Anderson-Darling
- Kolmogorov-Smirnov
- Cramer-von Mises

Data Distributions

Some Common Distributions

Continuous
- Normal
- Lognormal
- Inverse normal

Discrete
- Binomial
- Poisson

It is possible to analyze experiments with non-normal data (GLIMMIX), but needing to do so is rare.

Normalizing Data

Transformations

Lognormal data:

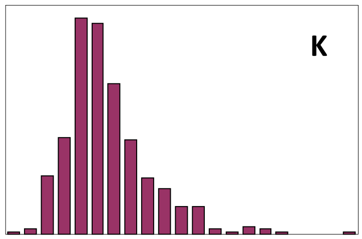
$$t_y = \log(y) \; or \; \ln(y)$$

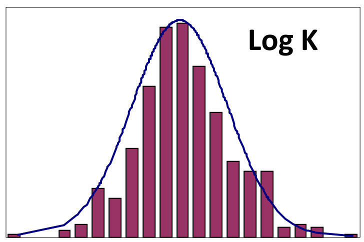
Left-skewed data:

$$t_y = y^2$$

Normalizing Data

Example – K Concentration

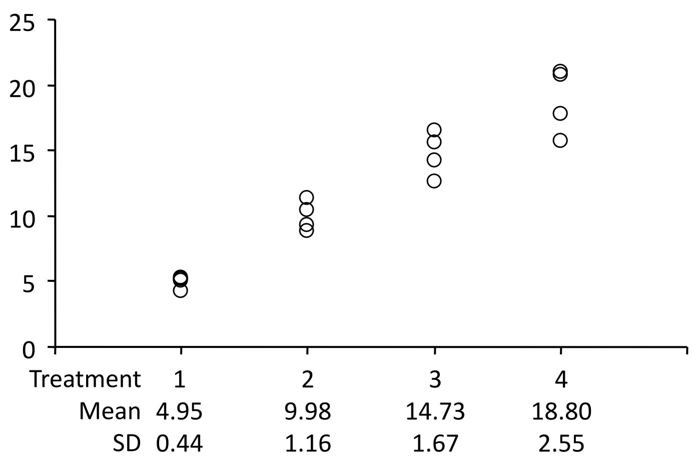




	K	Log K
Mean	26.583	1.404
SD	8.816	0.133
Median	24.860	1.396
Variance	77.714	0.018
Skew	1.495	0.266
Kurtosis	4.007	0.548

Heterogeneity

Unequal Variances



Heterogeneity
Common Issues

Certain data characteristics may lead to questions about homogeneity assumption:

- Whole numbers
- Percentages, proportions
- Very wide measurement response

Heterogeneity
Data Transformations

Condition	Transformation	Types of Data
$S \propto \bar{y}$	$\ln(y)$	growth data
$S \propto \bar{y}^2$	$\frac{1}{y}$	survival data, rate data
$S^2 \propto \bar{y}$	$\sqrt{y}$	whole number data, counts
$S^2 \propto 1-\bar{y}$	y^2	
$S^2 \propto \bar{y}(1-\bar{y})$	$\arcsin\sqrt{y}$	percentages, proportions

Tests for Homogeneity

- Bartlett's Test
- Hartley's F-max test
- Bartlett and Kendall ln s^2
- Modified Levine Test

Tests for Homogeneity

Rules for Assessing Significance

1. if accepted at α = .01
   → do not transform
2. if rejected at α = .001
   → transform
3. if result is between α = .01 and α = .001
   → transform only if there is a theoretical basis for doing so

Tests for Homogeneity

Bartlett's Test Example

Experiment	SS_e	df	S^2	$\ln S^2$
1	157.8	18	8.77	2.171
2	134.5	18	7.47	2.011
3	325.5	18	18.08	2.895
4	308.4	18	17.13	2.841
5	111.3	18	6.18	1.822
6	214.2	18	11.9	2.477
Total	1251.7	108	69.54	14.217

Tests for Homogeneity

Bartlett's Test Example

$$S_p^2 = \frac{1251.7}{108} = 11.59$$

$$\chi^2_{k-1} = \frac{(df)[k \ln s_p^2 - \sum_{i=1}^{k} \ln s_i^2]}{1 + (k+1)/[3(k)(df)]} = \frac{18[6(2.450) - 14.217]}{1 + (6+1)/[3(6)(18)]} = \frac{8.705}{1.022} = 8.521$$

$$\chi^2_{5,\,.001} = 20.52$$

$$\chi^2_{5,\,.01} = 15.09$$

$$\therefore accept\, H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2$$

Tests for Homogeneity

Hartley's F-max Test

$$F_{\max} = \frac{\max\left\{\hat{\sigma}_i^2\right\}}{\min\left\{\hat{\sigma}_i^2\right\}}$$

Numerator df = number of treatments, t

Denominator df = number of reps - 1, r - 1

Tests for Homogeneity

Hartley's F-max Test Example

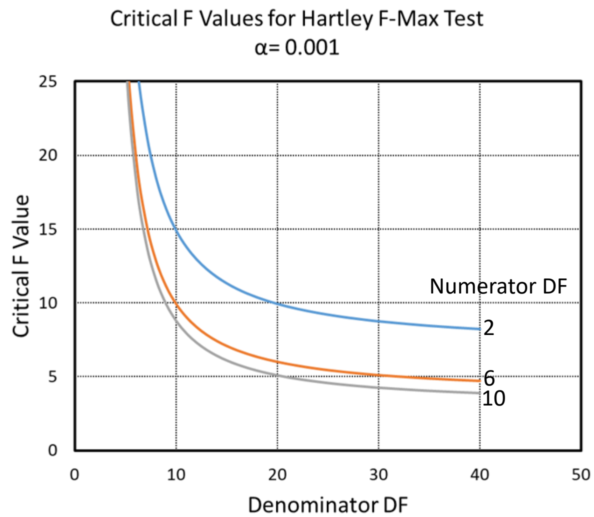
$$F_{\max} = \frac{\max\left\{\hat{\sigma}_i^2\right\}}{\min\left\{\hat{\sigma}_i^2\right\}} = \frac{18.08}{6.18} = 2.93$$

$F_{(.01,6,18)} = 4.01$

Therefore, do not reject H_0

Tests for Homogeneity

Hartley's F-max Test

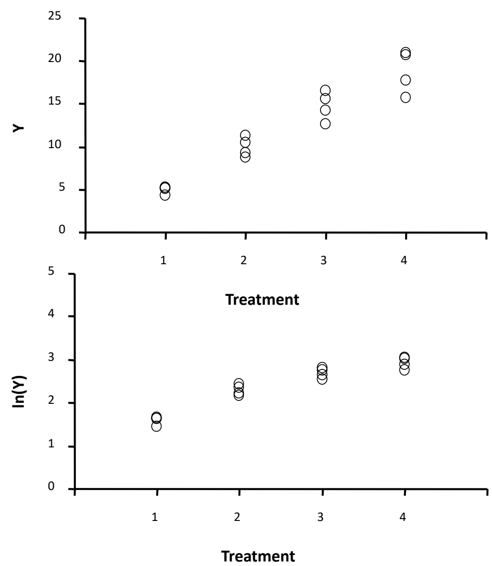


Stabilizing Variances

Transformation Example

Condition	Transformation	Types of Data
$S \propto \bar{y}$	$\ln(y)$	growth data
$S \propto \bar{y}^2$	$\frac{1}{y}$	survival data, rate data
$S^2 \propto \bar{y}$	$\sqrt{y}$	whole number data, counts
$S^2 \propto 1-\bar{y}$	y^2	
$S^2 \propto \bar{y}(1-\bar{y})$	$\arcsin\sqrt{y}$	percentages, proportions

Stabilizing Variances
Transformation Example

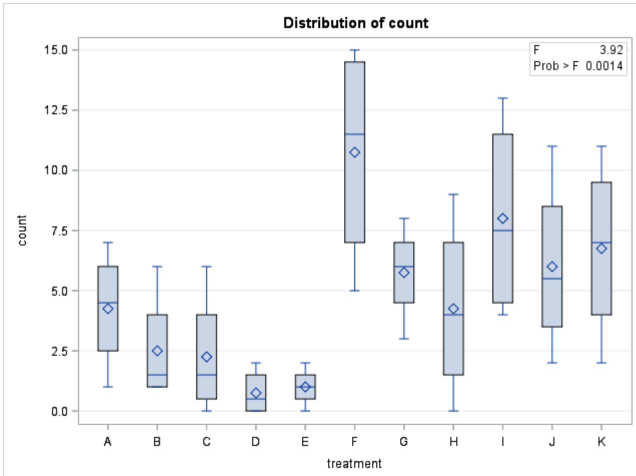


Stabilizing Variances
Transformation Example

	y				ln(y)			
Rep	1	2	3	4	1	2	3	4
1	4.28	10.47	16.49	17.77	1.45	2.35	2.80	2.88
2	5.16	8.77	15.55	20.74	1.64	2.17	2.74	3.03
3	5.10	11.35	14.23	15.67	1.63	2.43	2.66	2.75
4	5.25	9.31	12.66	21.02	1.66	2.23	2.54	3.05
Mean	4.95	9.98	14.73	18.80	1.60	2.30	2.68	2.93
s	0.45	1.16	1.67	2.55	0.09	0.12	0.12	0.14
s^2	0.20	1.35	2.77	6.52	0.01	0.01	0.01	0.02
CV	9.03	11.63	11.31	13.58	5.93	5.05	4.29	4.76

Transformation Example

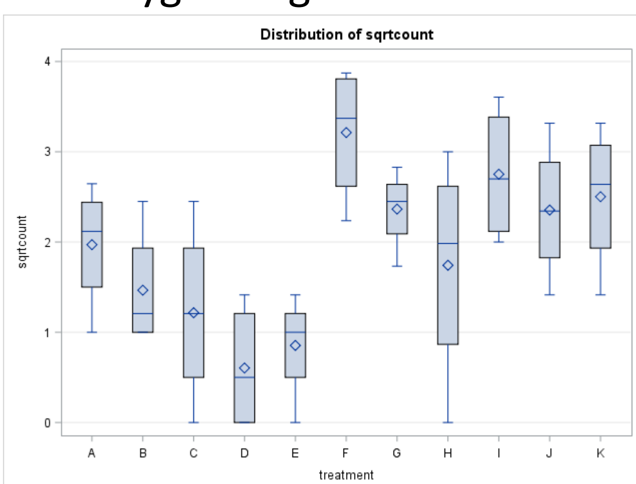
Lygus Bug Counts – Little and Hills



Treatment	Variance
A	6.2500
B	5.6667
C	6.9167
D	0.9167
E	0.6667
F	21.5833
G	4.2500
H	14.2500
I	18.0000
J	14.0000
K	14.2500

Transformation Example

Lygus Bug Counts – Little and Hills



Treatment	Variance
A	0.4897
B	0.4681
C	1.0287
D	0.5143
E	0.3619
F	0.5716
G	0.2099
H	1.6204
I	0.5762
J	0.6110
K	0.6521

Transformation Example

Lygus Bug Counts – Little and Hills

Treatment	t Grouping					Means		
						sqrt	original	inverse
F				A		3.21	10.75	10.32
I		B		A		2.75	8.00	7.57
K		B		A	C	2.50	6.75	6.26
G		B	D	A	C	2.36	6.00	5.54
J		B	D	A	C	2.35	5.75	5.59
A		B	D	E	C	1.97	4.25	3.88
H	F	B	D	E	C	1.74	4.25	3.03
B	F		D	E	C	1.47	2.50	2.15
C	F		D	E		1.22	2.25	1.48
E	F			E		0.85	1.00	0.73
D	F					0.60	0.75	0.36

Comparing Means

Unequal Variances

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{d}}{s_{\bar{d}}}$$

$$s_{\bar{d}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Comparing Means

Standard Error of a Difference

$$S_{\bar{d}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

if $s^2_1 = s^2_2$, then:

$$S_{\bar{d}} = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

if $s^2_1 = s^2_2$ and $n_1 = n_2$, then:

$$S_{\bar{d}} = \sqrt{\frac{2s_p^2}{n}}$$

Comparing Means

Unequal Variances

When $n_1 \neq n_2$:

$$df = \frac{\left(S_{\bar{x}_1}^2 + S_{\bar{x}_2}^2\right)^2}{\frac{\left(S_{\bar{x}_1}^2\right)^2}{n_1 - 1} + \frac{\left(S_{\bar{x}_2}^2\right)^2}{n_2 - 1}}$$

$$t'_{crit} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2} \qquad w_i = \frac{s_i^2}{n_i}$$

Comparing Means
PROC MIXED Example

PROC MIXED OUTPUT for Lab 4

Covariance Parameter Estimates

Cov Parm	Group	Estimate
Residual	type A	59.5556
Residual	type B	26.2222
Residual	type C	430.89
Residual	type D	436.22

Comparing Means
PROC MIXED Example

Least Squares Means

Effect	type	Estimate	Standard Error	DF	t Value	Pr > |t|
type	A	28.0000	2.4404	9	11.47	<.0001
type	B	13.0000	1.6193	9	8.03	<.0001
type	C	82.0000	6.5642	9	12.49	<.0001
type	D	154.00	6.6047	9	23.32	<.0001

For Type A:

$$S_{\bar{x}} = \sqrt{\frac{s^2}{r}} = \sqrt{\frac{59.56}{10}} = 2.44$$

Comparing Means
PROC MIXED Example

Differences of Least Squares Means

Effect	type	_type	Estimate	Standard Error	DF	t Value	Pr > |t|
type	A	B	15.0000	2.9288	15.6	5.12	0.0001
type	A	C	-54.0000	7.0032	11.4	-7.71	<.0001
type	A	D	-126.00	7.0411	11.4	-17.89	<.0001
type	B	C	-69.0000	6.7610	10.1	-10.21	<.0001
type	B	D	-141.00	6.8003	10.1	-20.73	<.0001
type	C	D	-72.0000	9.3119	18	-7.73	<.0001

For the comparison, A = B:

$$S_{\bar{d}} = \sqrt{\frac{59.56}{10} + \frac{26.22}{10}} = 2.93 \qquad df = \frac{(59.56 + 26.22)^2}{\frac{(59.56)^2}{10-1} + \frac{(26.22)^2}{10-1}} = 15.64$$