

STAT 5200 Handout #6: Checking Assumptions (Ch. 6)

Example: Researchers have 4 methods available to estimate the peak discharge in a watershed. They randomly select 24 comparable sites in the watershed and randomly assign 6 to each of the 4 methods. The estimates of peak discharge are obtained at each of the 24 sites, and the researchers want to see if the methods produce equivalent estimates of peak discharge.

CRD
w/ site
as exp. unit
? method
as factor (4 levels)

→ response var. y_i

```
/* Enter data */
data peak;
  input method peakEst @@;
  cards;
  1 0.34 1 0.12 1 1.23 ... 1 0.70 1 1.75 1 0.12
  2 0.91 2 2.94 2 2.14 2 2.36 2 2.86 2 4.55
  3 6.31 3 8.37 3 9.75 3 6.09 3 9.82 3 7.24
  4 17.15 4 11.82 4 10.95 4 17.20 4 14.35 4 16.82
  ;
```

→ read data continuously, so can have multiple observations per row

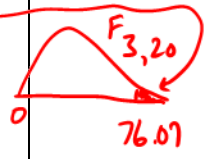
```
/* Fit CRD model and check assumptions */
proc glm data=peak plots=diagnostic;
  class method; (list factors)
  model peakEst = method;
  means method / HOVtest=Levene;
  title 'Comparison of all methods';
run;
```

test constant variance →

model $y =$ factors

→ homogeneity of variance

Comparison of all methods					
Dependent Variable: peakEst					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	708.3471125	236.1157042	76.07	<.0001
Error	20	62.0810833	3.1040542		
Corrected Total	23	770.4281958			



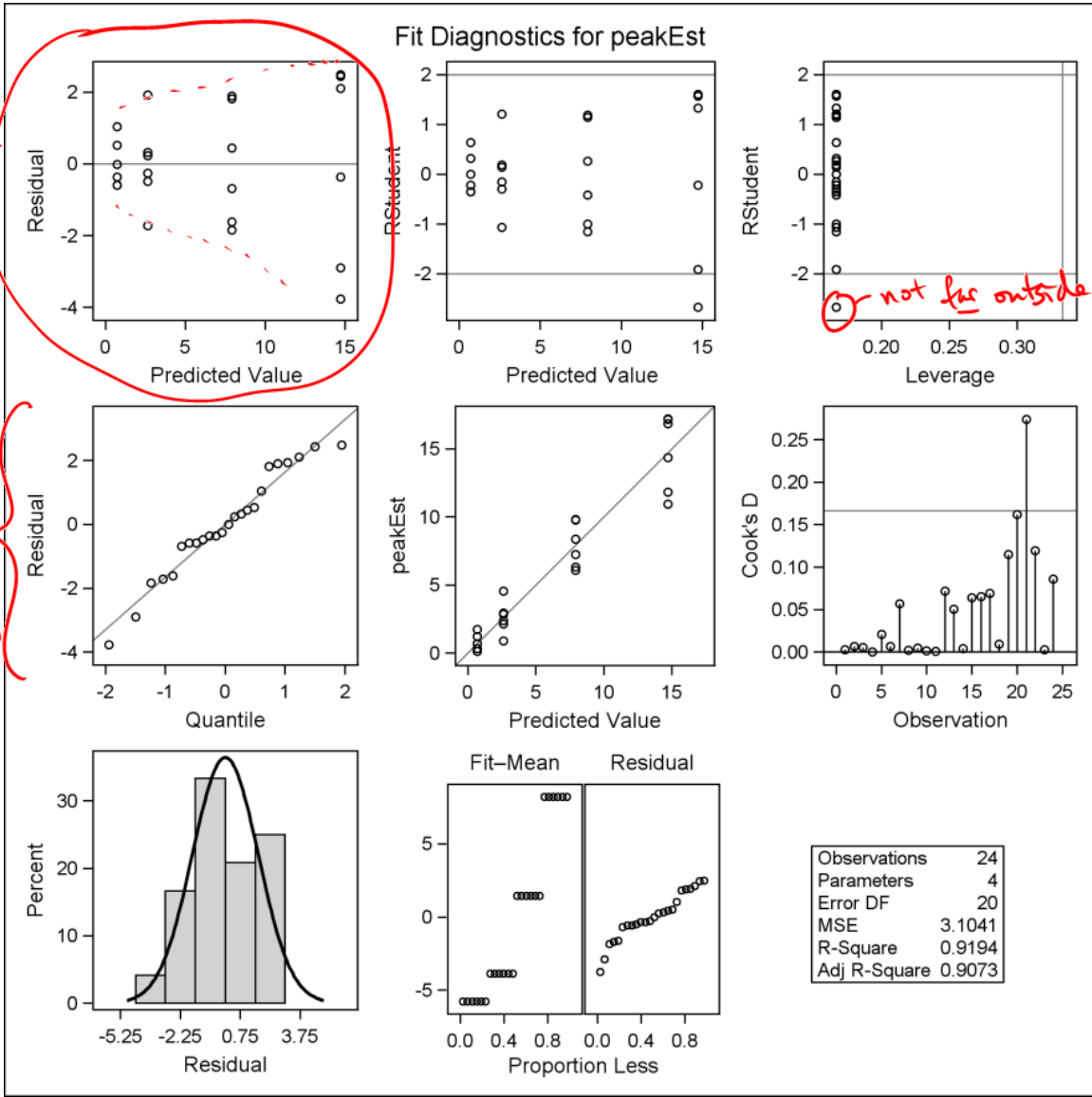
what is sampling distribution?
 $F_{3,20}$ or F with 3 numerator and 20 denom. DF

**Levene's Test for Homogeneity of peakEst Variance
ANOVA of Squared Deviations from Group Means**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
method	3	135.7	45.2417	6.75	0.0025
Error	20	134.1	6.7050		

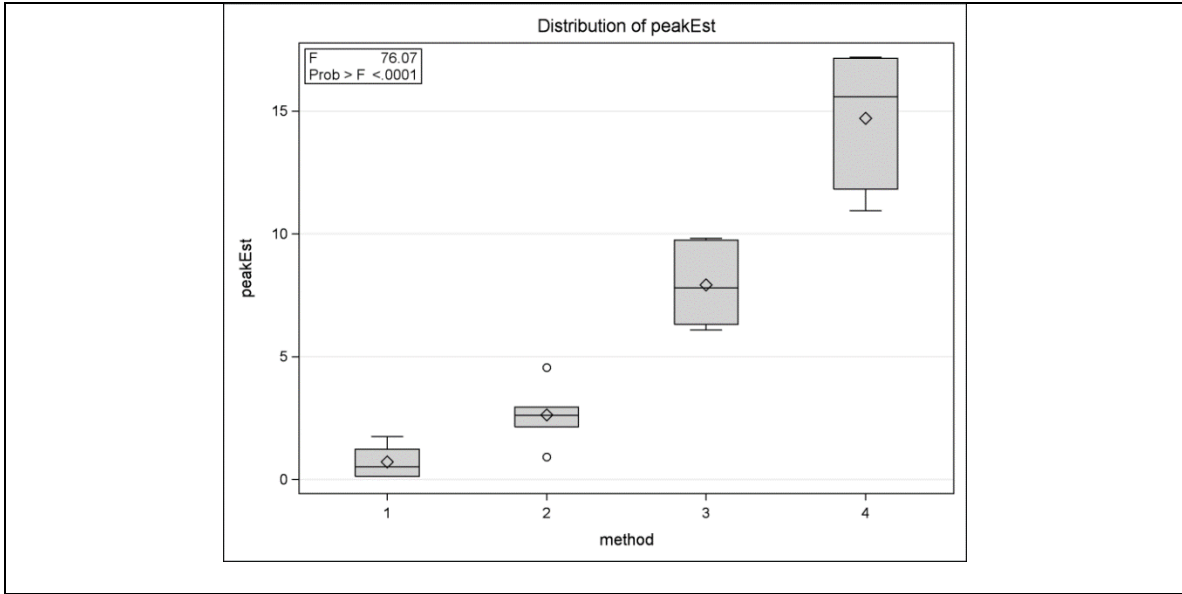
conclude evidence of non-constant variance

check constant variance



use to test for normality

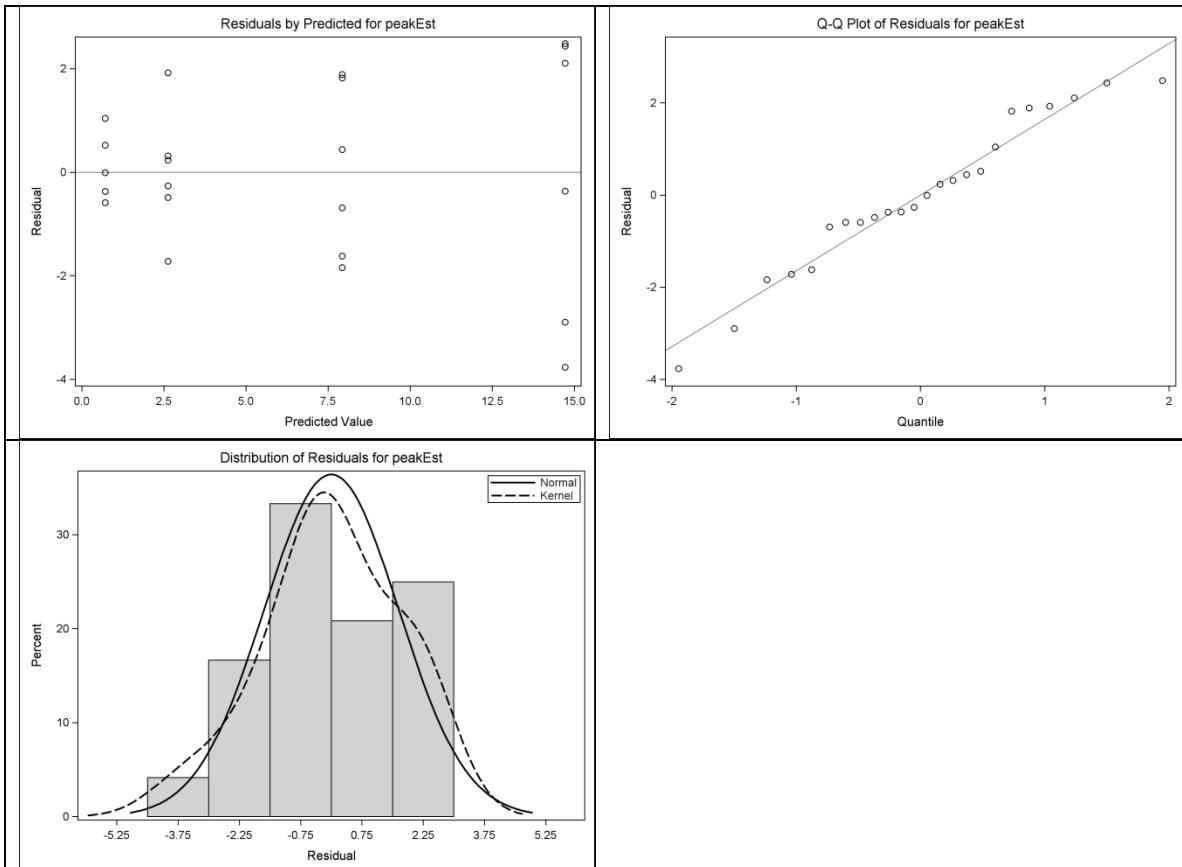
not far outside



```

/* Fit CRD model and check assumptions (focus on sub-plots)
*/
proc glm data=peak plots=diagnostic(unpack);
  class method;
  model peakEst = method;
run;

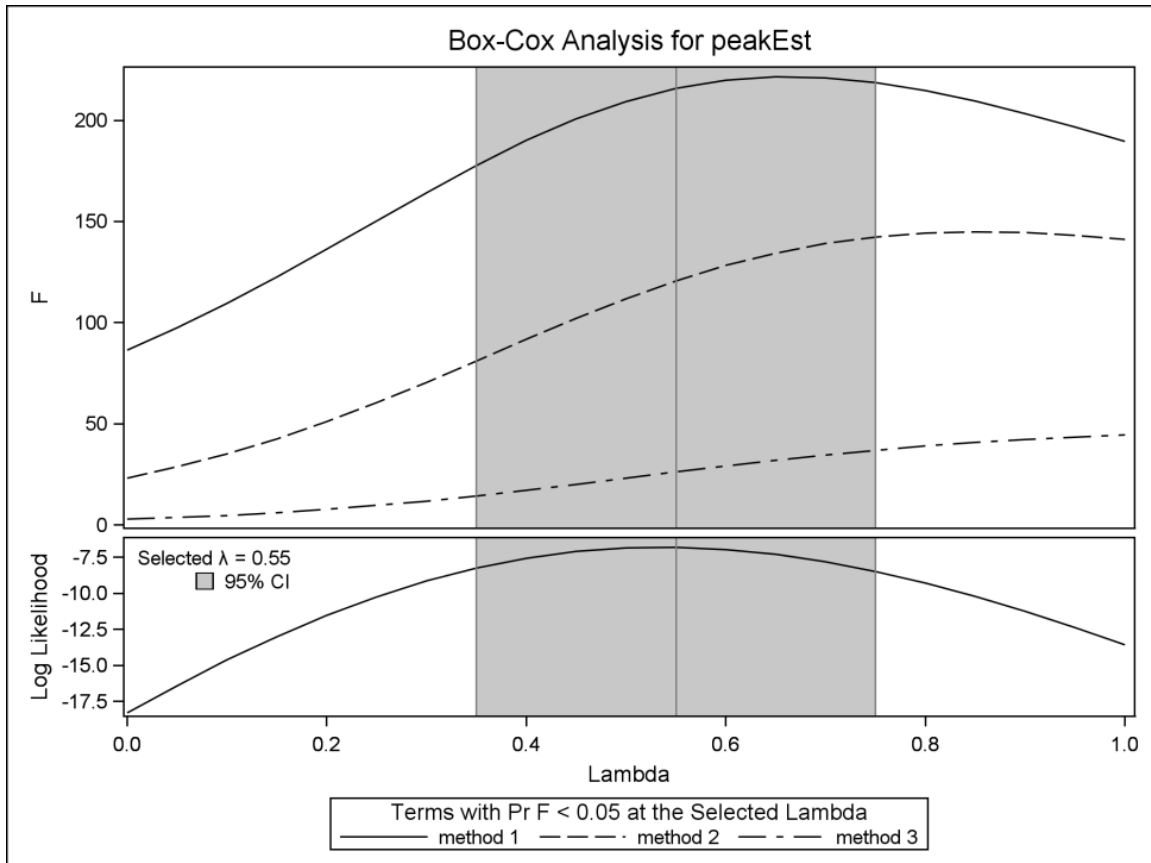
```



```

/* Consider transformation */
proc transreg data=peak;
  model boxcox(peakEst / lambda=0 to 1 by 0.05)
    = class(method);
  title 'Box-Cox on response';
run;

```



```

/* Re-fit model on transformed data */
data peak; set peak;
  sqrtPeakEst = peakEst**0.5;
run;

proc glm data=peak plots=diagnostic;
  class method;
  model sqrtPeakEst = method / solution;
  output out = out1 p=pred r=resid;
  means method / Bon HOVtest=Levene;
  title 'Comparison of all methods';
  title2 '(square root scale)';
run;

```

**Comparison of all methods
(square root scale)**

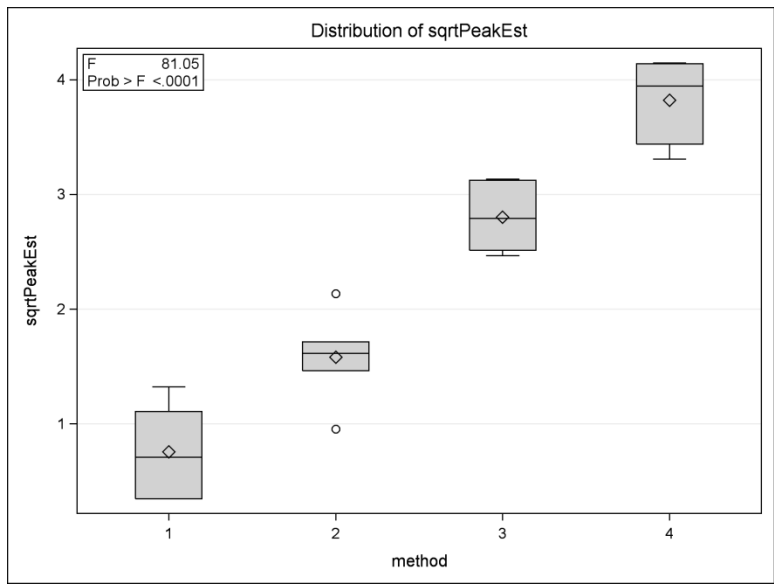
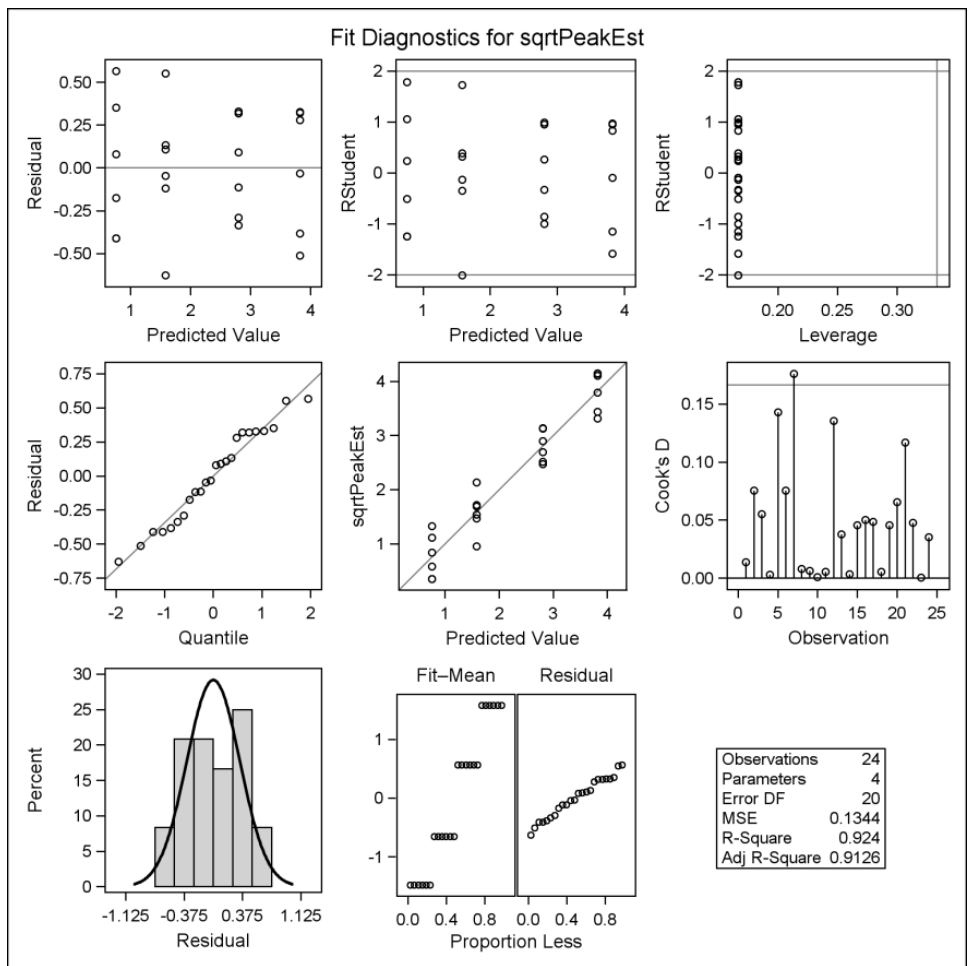
The GLM Procedure

Dependent Variable: sqrtPeakEst

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	32.68421267	10.89473756	81.05	<.0001
Error	20	2.68843280	0.13442164		
Corrected Total	23	35.37264547			

R-Square	Coeff Var	Root MSE	sqrtPeakEst Mean
0.923997	16.36121	0.366636	2.240883

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	3.820833991	B	0.14967834	25.53	<.0001
method 1	-3.063416516	B	0.21167714	-14.47	<.0001
method 2	-2.238848756	B	0.21167714	-10.58	<.0001
method 3	-1.017538852	B	0.21167714	-4.81	0.0001
method 4	0.000000000	B	.	.	.



Levene's Test for Homogeneity of sqrtPeakEst Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
method	3	0.0143	0.00478	0.36	0.7858
Error	20	0.2691	0.0135		

Bonferroni (Dunn) t Tests for sqrtPeakEst

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	20
Error Mean Square	0.134422
Critical Value of t	2.92712
Minimum Significant Difference	0.6196

Means with the same letter are not significantly different.			
Bon Grouping	Mean	N	method
A	3.8208	6	4
B	2.8033	6	3
C	1.5820	6	2
D	0.7574	6	1