Stat 541 Homework #4

Kenny Flagg

February 24, 2016

1. (a) **ANOVA on Untransformed Data**

H_0: $\mu_1^2 = \mu_2^2 = \mu_3^2 = \mu_4^2 = \mu_5^2$; The mean count per second is the same for each treatment.
H_a: $\mu_i^2 = \mu_j^2$ for some $i \neq j$; Some treatment means are not equal.

```
                  ANOVA RESULTS: No Transformation

                         The GLM Procedure

                  Dependent Variable: countsec

                                      Sum of
Source                       DF      Squares   Mean Square  F Value  Pr > F

Model                         4   0.04752280    0.01188070    95.53  <.0001

Error                        25   0.00310920    0.00012437

Corrected Total              29   0.05063200


          R-Square     Coeff Var      Root MSE    countsec Mean

          0.938592      22.30408      0.011152         0.050000


Source                       DF  Type III SS   Mean Square  F Value  Pr > F

pCig                          4   0.04752280    0.01188070    95.53  <.0001
```

The `pCig` row has p-value < 0.0001 so there is strong evidence that the treatments do not all have the same mean count per second.

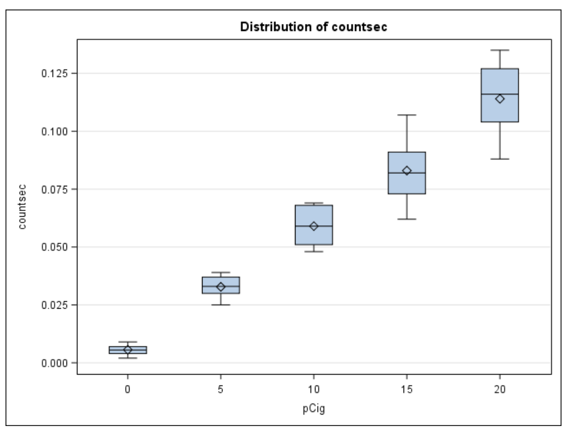
(b) **Tests of Contrasts**

```
                    ANOVA RESULTS: No Transformation

                           The GLM Procedure

                     Dependent Variable: countsec
```

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Linear	0.26700000	0.01410634	18.93	<.0001
Quadratic	0.00540000	0.01727667	0.31	0.7572
Cubic	0.00800000	0.01537203	0.52	0.6073
Quartic	0.01040000	0.04157782	0.25	0.8045

The linear contrast has p-value < 0.0001, so there is strong evidence of a linear trend in the treatment mean count per second. The other contrasts all have p-value > 0.05 so there is little to no evidence of quadratic, cubic, or quartic trends.

(c) **Boxplots**



The results from the tests of the contrasts agree with the boxplots, which show a very linear trend with no curvature.

(d) **Levene's Test**

H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$; The variances are equal for each treatment.
H_a: $\sigma_i^2 = \sigma_j^2$ for some $i \neq j$; Some treatment variances are not equal.

ANOVA RESULTS: No Transformation

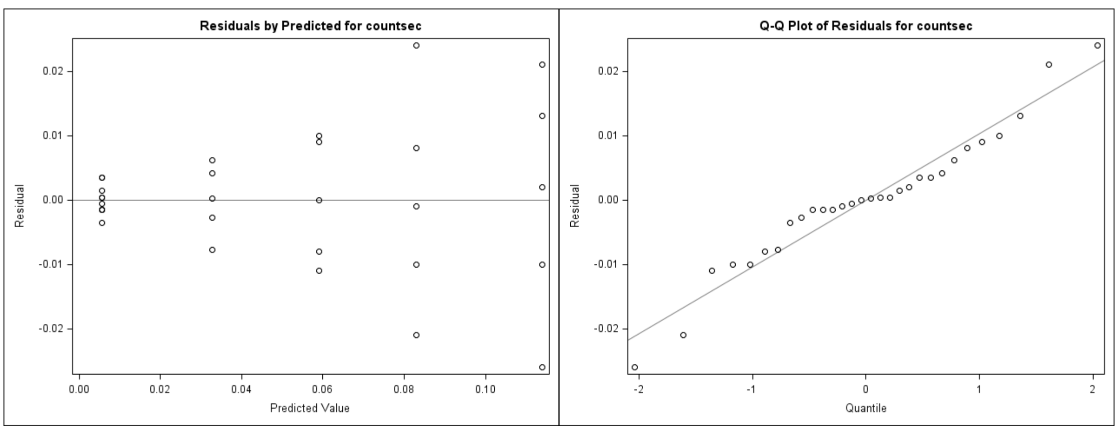
The GLM Procedure

Levene's Test for Homogeneity of countsec Variance
ANOVA of Absolute Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
pCig	4	0.000752	0.000188	5.58	0.0024
Error	25	0.000842	0.000034		

Levene's test gives a p-value = 0.0024 so there is strong evidence that the variances differ between treatments.

(e) **Diagnostic Plots**



The residuals vs predicted values plot shows a clear fanning pattern of increasing variance with increasing counts per second. The normal probability plot does not show any serious deviation from normality aside from some extreme values in the tails. It is appropriate to use Levene's test to check the homogeneity of variance assumption.

(f) **ANOVA on Square-Root Transformed Response**

H_0: $\mu_1^2 = \mu_2^2 = \mu_3^2 = \mu_4^2 = \mu_5^2$; The mean square root count per second is the same for each treatment.

H_a: $\mu_i^2 = \mu_j^2$ for some $i \neq j$; Some treatment means are not equal.

```
                    ANOVA RESULTS: Square Root Transformation

                              The GLM Procedure

                        Dependent Variable: sqrtcnt

                                        Sum of
Source                      DF         Squares    Mean Square   F Value   Pr > F

Model                        4      0.30234533     0.07558633    164.99   <.0001

Error                       25      0.01145339     0.00045814

Corrected Total             29      0.31379872


          R-Square     Coeff Var      Root MSE    sqrtcnt Mean

          0.963501      10.76412      0.021404        0.198847


Source                      DF     Type III SS    Mean Square   F Value   Pr > F

pCig                         4      0.30234533     0.07558633    164.99   <.0001
```

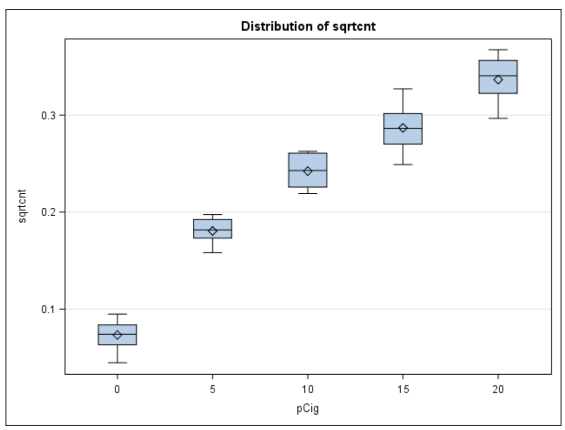
The `pCig` row has p-value < 0.0001 so there is strong evidence that the treatments do not all have the same mean square root count per second.

Tests of Contrasts

```
                ANOVA RESULTS: Square Root Transformation

                            The GLM Procedure

                       Dependent Variable: sqrtcnt

                                               Standard
Parameter                      Estimate           Error    t Value    Pr > |t|

Linear                       0.63300870      0.02707429      23.38      <.0001
Quadratic                   -0.13182571      0.03315909      -3.98      0.0005
Cubic                        0.05075728      0.02950352       1.72      0.0977
Quartic                     -0.00612794      0.07980028      -0.08      0.9394
```

The linear contrast has p-value < 0.0001 and the quadratic contrast has p-value $=$ 0.0005, so there is strong evidence of linear and quadratic trends in the mean square root count per second. The cubic and quartic contrasts have p-value > 0.05 so there is little to no evidence of cubic or quartic trends in the transformed data.

Boxplots



The boxplots show an increasing trend with some quadratic curvature, but no higher-order curvature is apparent. This agrees with the results from testing the contrasts.

Levene's Test

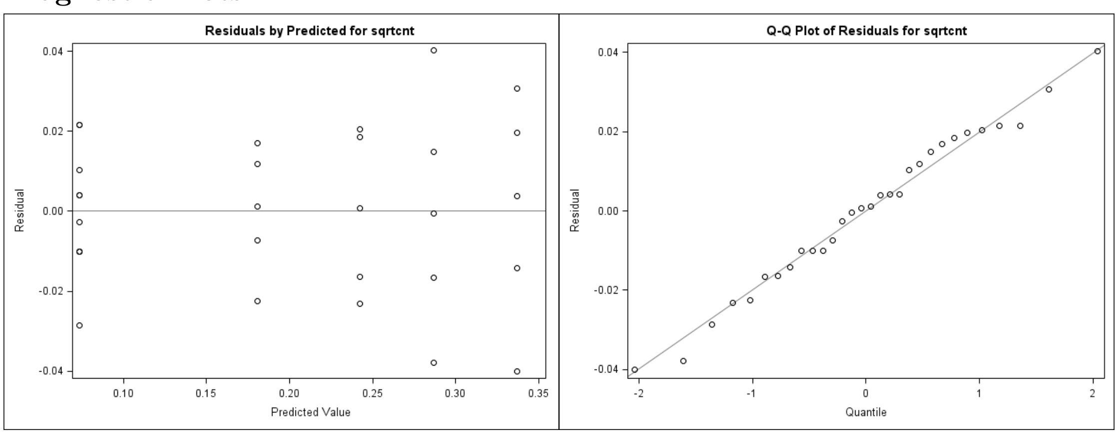
H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$; The variances are equal for each treatment.
H_a: $\sigma_i^2 = \sigma_j^2$ for some $i \neq j$; Some treatment variances are not equal.

```
                ANOVA RESULTS: Square Root Transformation

                            The GLM Procedure

            Levene's Test for Homogeneity of sqrtcnt Variance
              ANOVA of Absolute Deviations from Group Means
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
pCig	4	0.000563	0.000141	1.10	0.3773
Error	25	0.00319	0.000128		

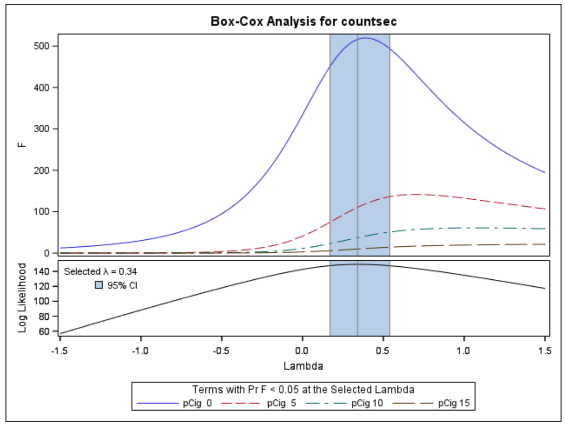
Levene's test on the transformed data gives p-value = 0.3773. There is little to no evidence that that variances differ between the treatments. The transformation has stabilized the variance.

Diagnostic Plots



The residuals vs predicted values plot shows very similar spreads for each predicted value, so the variance can be assumed constant. The normal probability plot shows no problems.

(g) **Box-Cox Transformation**



The Box-Cox method recommends $\lambda = 0.34$, or approximately $y^* = \sqrt[3]{y}$.

(h) **Recommendation**

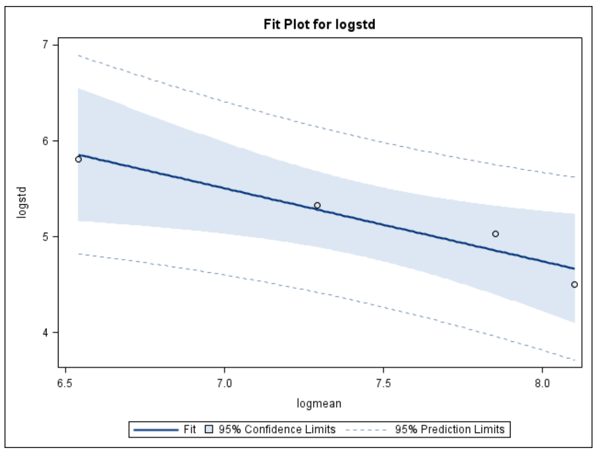
I recommend using the square root transformation. Levene's test on the original data gives evidence (p-value = 0.0024) the the homogeneity of variance assumption does not hold, so ANOVA results on the original data are unreliable. The square root transformation is reasonable because 0.5 is in the 95% confidence interval for λ from the Box-Cox procedure.

2. (a) The boxplot shows vastly different spreads among the groups, so yes it was reasonable to do a WLS ANOVA using $w_i = \dfrac{1}{s_i^2}$ as the weights.

   (b) The spread of a group does not appear to be related to the mean of the group, so the Box-Cox procedure is unlikely to successfully stabilize the variance.

3. Fit the regression model $\log(s_i) = \log(\theta) + \alpha \log(\bar{y}_i) + \epsilon_i$

```
                 ANOVA TO FIND EMPIRICAL SELECTION OF ALPHA

                            The GLM Procedure

                       Dependent Variable: logstd

                                         Standard
Parameter           Estimate                Error    t Value    Pr > |t|

Intercept        10.83673889           1.10964799       9.77      0.0103
logmean          -0.76145545           0.14854676      -5.13      0.0360
```



The least squares estimate of the slope is $\widehat{\alpha} = -0.7614$ so the suggested transformation is $y^* = y^{1.7614}$

Appendix: SAS Code

Problem 1

```
OPTIONS CENTER NODATE NONUMBER LINESIZE=75 FORMDLIM='*';
ODS LISTING FILE='hw4p1.lis';

DATA in;
INPUT pCig n @@;
DO i = 1 to n;
  INPUT countsec @@;
  sqrtcnt = sqrt(countsec);
  OUTPUT;
END;
DATALINES;
 0 10 .004 .007 .004 .009 .009 .006 .004 .006 .002 .005
 5  5 .030 .025 .037 .039 .033
10  5 .069 .068 .048 .059 .051
15  5 .082 .107 .062 .073 .091
20  5 .135 .104 .088 .127 .116
;


/*******************************************
 *** ANOVA on the untransformed responses ***
 *******************************************/

PROC GLM DATA=in PLOTS(UNPACK)=ALL;
  ODS GRAPHICS / IMAGENAME='orig' RESET=INDEX;
  CLASS pCig;
  MODEL countsec = pCig / SS3;
  MEANS pCig / HOVTEST=LEVENE(TYPE=ABS);
  ESTIMATE 'Linear   ' pCig -2 -1  0  1  2;
  ESTIMATE 'Quadratic' pCig  2 -1 -2 -1  2;
  ESTIMATE 'Cubic    ' pCig -1  2  0 -2  1;
  ESTIMATE 'Quartic  ' pCig  1 -4  6 -4  1;
TITLE 'ANOVA RESULTS: No Transformation';


/***********************************************
 *** Use a default square root transformation ***
 ***********************************************/

PROC GLM DATA=in PLOTS(UNPACK)=ALL;
  ODS GRAPHICS / IMAGENAME='sqrt' RESET=INDEX;
  CLASS pCig;
```

```
  MODEL sqrtcnt = pCig / SS3;
  MEANS pCig / HOVTEST=LEVENE(TYPE=ABS);
  ESTIMATE 'Linear   ' pCig -2 -1  0  1  2;
  ESTIMATE 'Quadratic' pCig  2 -1 -2 -1  2;
  ESTIMATE 'Cubic    ' pCig -1  2  0 -2  1;
  ESTIMATE 'Quartic  ' pCig  1 -4  6 -4  1;
TITLE 'ANOVA RESULTS: Square Root Transformation';


/*******************************************************
 *** Find the transformation using the Box-Cox method ***
 *******************************************************/

PROC TRANSREG DATA=in;
  ODS GRAPHICS / IMAGENAME='boxcox' RESET=INDEX;
  MODEL BOXCOX(countsec / LAMBDA=-1.5 to 1.5 by .01) = CLASS(pCig);
TITLE 'Find the Box-Cox Transformation using PROC TRANSREG';

run;
quit;
```

Problem 3

```
options center nodate nonumber linesize=75 formdlim='*';
ods listing file='hw4p3.lis';

DATA in;
input nmole mean std;
datalines;
  0  3296   90
  1  2574  153
 10  1466  207
100   692  332
;

DATA yset;
  SET in;
  logstd = LOG(std);
  logmean = LOG(mean);

PROC GLM DATA=yset;
  ODS GRAPHICS / IMAGENAME='empirical' RESET=INDEX;
  MODEL logstd = logmean / SS3 solution;
TITLE 'ANOVA TO FIND EMPIRICAL SELECTION OF ALPHA';

run;quit;
```