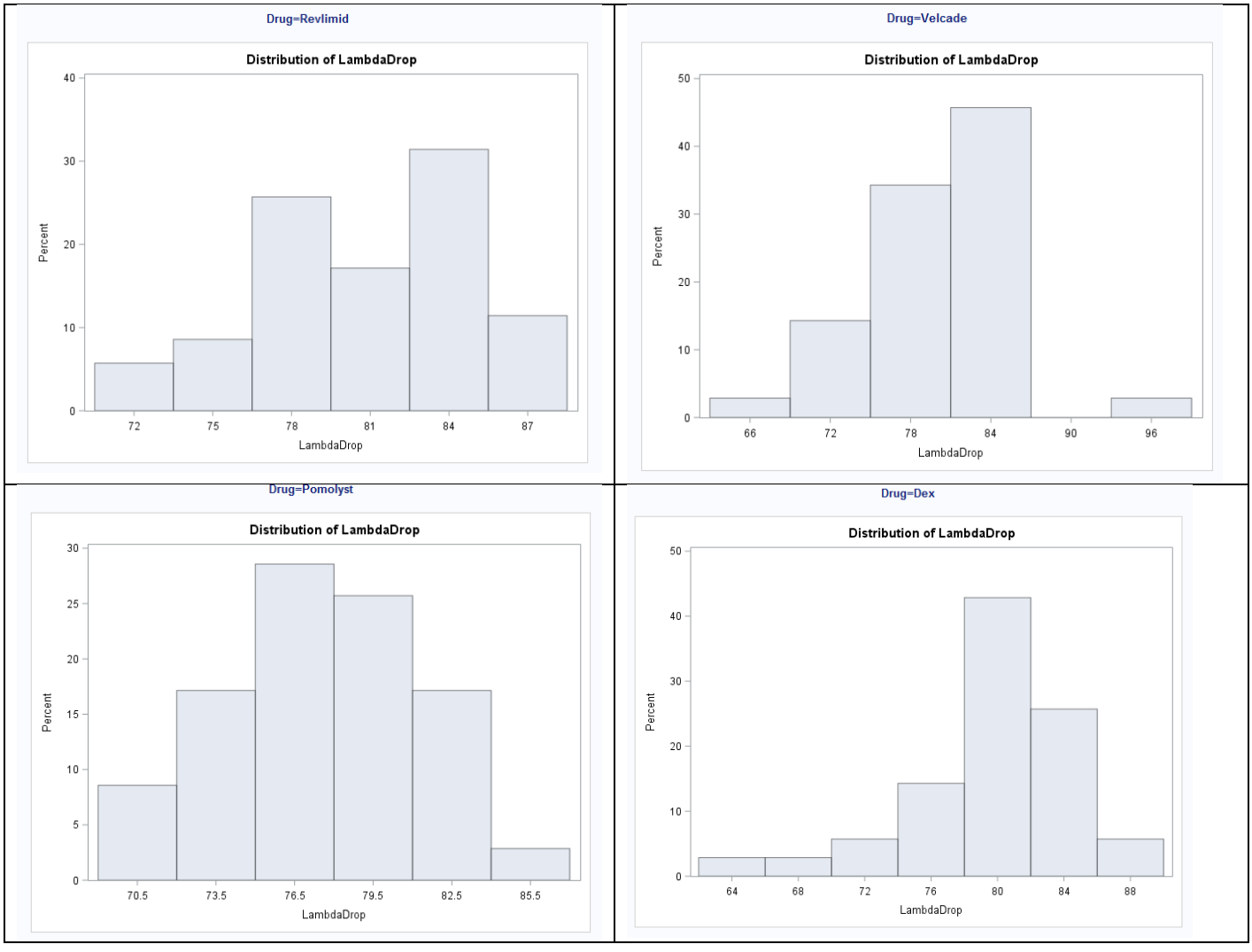
MSDS 6371 Analysis Question

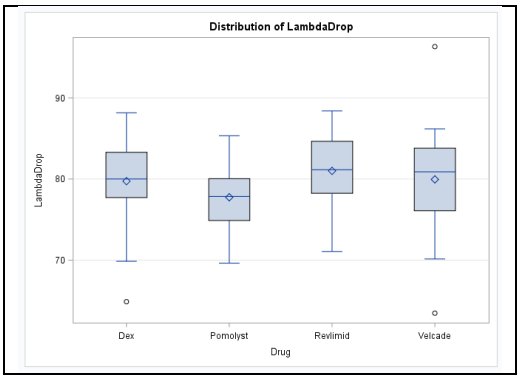
Multiple Myeloma is a relatively rare form of blood cancer in which plasma cells reproduce uncontrollably in the bone marrow. This in turn crowds out read and white blood cells which leads to complications in the patients overall health. The cancer can be monitored by looking at a patient's *Lambda and Kappa* protein counts. Higher counts indicate more / worse disease. There have been several drugs that have been developed to fight this type of cancer which is measured in the reduction of the *Lamda* and *Kappa* protein levels. Four of these drugs are: Revlimid, Velcade, Dex and Pomolyst.

Through prior studies, all of the drugs have been shown to decrease the Lamda and Kappa proteins in myeloma patients. This study is focused on comparing the four treatments to see which ones are more effective in reducing the Lamda and Kappa protein levels.

To test this, the researchers went to MD Anderson hospital in Houston and randomly selected 35 patient's records that had been taking one of the drugs above. To be clear, there were 140 patient's records in the study, 35 patients that took Revlimid, 35 that took Velcade, 35 that took Dex, and 35 that took Pomolyst. The researcher recorded the percentage drop in Lamda protein during the standard 15 month treatment of taking the drug. This data is recorded in the file myeloma.csv on canvas.

Histograms and Box Plots of the percent drops of **Lambda** protein levels are provided below.





Make sure and provide your SAS or R code for all questions below.

a. The principal question of interest for the researchers was if there were significant differences in the mean or median percent drop of Lambda protein between the drugs and if there are, and estimate of the magnitude of the difference(s). Provide analysis that will best answer the principle question of interest above. Please address all assumptions needed to conduct your analysis and provide a scope of inference with your findings. Assume the researchers are interested in maintaining a family wise error rate of 5% (alpha_family = .05).

Normality: Judging from the histograms, QQ plots, and box plots, distribution of percent drop of Lambda proteins of all four drugs are close to normal. Velcade has left skew with one outlier on each side but that does not preclude from assumption of normality. Additionally, all the data sets have same number of sample sizes.

Level of Drug	N	LambdaDrop	
		Mean	Std Dev
Dex	35	79.7293748	4.85470096
Pomolyst	35	77.7525305	3.74755333
Revlimid	35	81.0053776	4.24741071
Velcade	35	79.9517750	5.79332337

Homogeneity of standard deviations: Standard deviations of all data sets are fairly similar

Independence of Observations: Coming from mutually exclusive groups these populations are independent of each other group-wise. The study states that these are random samples from patients' charts. Data comes from available patient, this may constrain from making inference to general population difficult. In general we can answer the causation.

With the assumptions that data distributions of percent drop of Lambda proteins for all four drugs are normal with fairly equal standard deviations we will proceed with one-way ANOVA test on given data.

Aim: Test if any drug has effect in terms of percent drop in Lambda protein distribution different than the others.

- $Reduced\ model: \mu_{<D} = \mu_P = \mu_R = \mu_V$
- $Full\ model: at\ least\ one\ pair\ \neq$

The GLM Procedure
Dependent Variable: LambdaDrop

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	193.491612	64.497204	2.89	0.0377
Error	136	3033.322605	22.303843		
Corrected Total	139	3226.814218			

R-Square	Coeff Var	Root MSE	LambdaDrop Mean
0.059964	5.932305	4.722694	79.60976

Conclusion: There is moderate evidence that at least one of the drugs has an effect that is different than others ($p < 0.05$ from one-way ANOVA on log transformed data)

Aim2: Test the evidence of specific difference in mean drug effect in terms of percent drop of Lambda proteins for different drugs

With the assumptions of normality and homogeneity of standard deviations of data drug effect in terms of percent drop of Lambda proteins for different drugs and moderate and equal number of sample per group we can proceed with Tukey-Kramer procedure for adjustment of multiplication factor to compute confidence interval for individual groups mean differences

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

Drug	LambdaDrop LSMEAN	LSMEAN Number
Dex	79.7293748	1
Pomolyst	77.7525305	2
Revlimid	81.0053776	3
Velcade	79.9517750	4

Least Squares Means for effect Drug
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: LambdaDrop

i/j	1	2	3	4
1		0.3016	0.6715	0.9973
2	0.3016		0.0236	0.2131
3	0.6715	0.0236		0.7870
4	0.9973	0.2131	0.7870	

Alpha	0.05
Error Degrees of Freedom	136
Error Mean Square	22.30384
Critical Value of Studentized Range	3.67848
Minimum Significant Difference	2.9365

Comparisons significant at the 0.05 level are indicated by ***.

Drug Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Revlimid - Velcade	1.054	-1.883	3.990	
Revlimid - Dex	1.276	-1.660	4.212	
Revlimid - Pomolyst	3.253	0.316	6.189	***
Velcade - Revlimid	-1.054	-3.990	1.883	
Velcade - Dex	0.222	-2.714	3.159	
Velcade - Pomolyst	2.199	-0.737	5.136	
Dex - Revlimid	-1.276	-4.212	1.660	
Dex - Velcade	-0.222	-3.159	2.714	
Dex - Pomolyst	1.977	-0.960	4.913	
Pomolyst - Revlimid	-3.253	-6.189	-0.316	***
Pomolyst - Velcade	-2.199	-5.136	0.737	
Pomolyst - Dex	-1.977	-4.913	0.960	

Conclusion2: There is moderate evidence at significance level alpha < 0.05 (p = 0.0236) that mean effect level is different for Pomlyst compared to Revlimid. Estimate of the magnitude of the difference(s) between drugs are as below.

Drug	LambdaDrop LSMEAN	95% Confidence Limits	
Dex	79.729375	78.150725	81.308024
Pomolyst	77.752530	76.173881	79.331180
Revlimid	81.005378	79.426728	82.584027
Velcade	79.951775	78.373126	81.530424

Least Squares Means for Effect Drug				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	1.976844	-0.959619	4.913308
1	3	-1.276003	-4.212466	1.660461
1	4	-0.222400	-3.158864	2.714063
2	3	-3.252847	-6.189310	-0.316384
2	4	-2.199244	-5.135708	0.737219
3	4	1.053603	-1.882861	3.990066

Scope of inference: The data was acquired with random sample from available patient reports thus we one can claim that Revmild cause significantly more percent drop in Lambda protein than Pomolyst.

b. A second question of interest centered on comparing Type A drugs to Type B drugs. Turns out that Revlimid and Velcade are Type A drugs while Dex and Pomolyst are both Type B drugs. Test the claim that the Type A drugs have a greater mean percent drop of Lambda protein than the Type B drugs by comparing the mean percent drop of Revlimid and Velcade to the mean percent drop of Dex and Pomolyst using a contrast. For this question you may assume all the assumptions are met to run a contrast but do show all 6 steps of the hypothesis test (t test) and provide a 95% confidence interval for the difference as well. Test at the alpha = .01 level of significance.

With the assumptions that data distributions of percent drop of Lambda proteins for all four drugs are normal with fairly equal standard deviations we will proceed with one-way ANOVA test on given data.

Aim: Test the claim that the Type A drugs have a greater mean percent drop of Lambda protein than the Type B drugs

Separate Revlimid and Velcade (A) and Dex and Pomolyst (B) in two groups to create a contrast

- $Reduced\ model: \mu_{<R\_V} = \mu_{D\_P}$
- $Full\ model: at\ least\ one\ pair\ \neq$

t_critical = 2.353 for alpha 0.01

t_critical = 1.656for alpha 0.05

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Estimate A to B	1.73762364	0.79828106	2.18	0.0312

T value = 2.18

P value 0.0312 > 0.01

On the basis of this test, there is not enough evidence at significance level of 0.01 to reject the claim that the mean effect of type A and type B drugs are the same (pvalue = .0312 from a two sided t-test). There is moderate evidence to suggest that at type A drugs are more effective than type B drugs with difference in mean percent drop in Lambda protein level for type A drug being 1.74 percent higher than type B with 95% confidence interval of 3.06 to 0.416 percent. The data was acquired with random sample from available patient reports thus we one can claim that there is an indication that type A drugs cause more percent Lambda protein drop than the type B and further studies need to be conducted.

```
proc IMPORT OUT = WORK.DRUGoutcome
            DATAFILE="/home/mpednekar0/BridgeCourse/myeloma.csv"
            DBMS=csv REPLACE;
        GETNAMES=YES;
        DATAROW=2;
RUN;

/* Check stats, histograms, QQplots */
proc univariate data=DRUGoutcome;
class Drug;
    var LambdaDrop;
    histogram LambdaDrop;
    qqplot LambdaDrop;
run;
/* Test equality of Std Dev of income */
proc glm data=DRUGoutcome;
class Drug;
model LambdaDrop = Drug;
means Drug / HOVTEST=BF;
run;


/* Test group-wise diff of drugs */
proc glm data=DRUGoutcome;
class Drug;
model LambdaDrop = Drug;
means Drug / HOVTEST=BF tukey cldiff;
lsmeans Drug / pdiff adjust = tukey cl;
run;


proc glm data=DRUGoutcome order=data;
class Drug;
model LambdaDrop = Drug;
estimate 'Estimate A to B' Drug 0.5 0.5 -0.5 -0.5;
run;
```

c. BONUS (5 pts) **KAPPA** PROTEIN ANALYSIS: The researchers thought that Revlimid would be more effective than the Velcade in reducing the **Kappa** protein levels. One problem they encountered was that the Kappa percent reduction was missing for many of the Revlimid and Velcade patients (as can be seen in the data set). Conduct a complete analysis (state the problem, address the assumptions, conduct the 6 step test and provide a scope of inference) that will test the claim that the Revlimid has a greater mean or median percent drop in **Kappa** protein levels than Velcade. Use an alpha = .01 level of significance and provide confidence intervals with your analysis.

Normality: Judging from the histograms, QQ plots, and box plots, all distributions for Kappa results are right skewed and excessively asymmetric. All groups have outliers drug R has to many outliers for sample size of 10. Additionally, all the data sets have different number of sample sizes, R and V have very few (10) samples.

Given the small number of samples and non-normal data plus small number of samples we will run Wilcoxon sign rank test

```
/* Check stats, histograms, QQplots */
proc univariate data=DRUGoutcome;
class Drug;
    var KappaDrop;
    histogram KappaDrop;
    qqplot KappaDrop;
run;

/* Test equality of Std Dev of income */
proc glm data=DRUGoutcome;
class Drug;
model KappaDrop = Drug;
means Drug / HOVTEST=BF;
run;

proc glm data=DRUGoutcome order=data;
class Drug;
model KappaDrop = Drug;
estimate 'Estimate R to V' Drug 0 0 1 -1;
run;


ods output WilcoxonTest = obsnpar;
proc npar1way data = DRUGoutcome wilcoxon;
var KappaDrop;
class Drug;
exact wilcoxon;
run;
```