BIOSTAT 200B HW7

1.

Q.1: How do we interpret the parameter estimates in this output? Specifically, how do these parameter estimates relate to the group means?

GLM procedure here creates three dummy variables, one for each level of Drug. Drug G is the reference level and has a parameter estimate of 0. The parameter estimates for other levels of Drug represent the difference in the effect of each level compared with the reference level, Drug G.

The Intercept (62.1) estimates the mean of Y for Drug G.

The Drug A estimate (-37.8) is the estimated difference in the mean of Y between Drugs A and G, and the Drug D estimate (-20.9) is the estimated difference in the mean of Y between Drugs D and G.

Q.2: Where in the plots do you see evidence of violation of the homogeneity of variance assumption?

We can see the evidence of violation of the homogeneity of variance assumption in residual plot, studentized residual plot, and the residuals vs leverage plot. All of the spread of standardized residuals seem to change as a function of predicted value or leverage: here them appear to decrease, indicating heteroscedasticity. Also, note that the plot of Cook's D statistic indicates that observations in the higher, more variable age groups are overly influential on the analysis of group means. The overall inference from these plots is that an assumption of equal group variances is probably untenable and that the analysis of the group means should thus take this into account.

Q.3: Write code and estimate the pairwise contrasts between the youngest age group and each other age group. Use an appropriate method to control for multiple comparisons.

```
means agegroup / dunnettl;
run;
```

The multiple comparison results are shown below. Since we want to know the pairwise contrasts between the youngest age group (agegroup = 1) and each other group (agegroup = 2, 3, 4, 5), we use Dunnett's method and let the youngest age group be the control group, while letting other groups be the treatment groups.

The GLM Procedure

Dunnett's One-tailed t Tests for smell

Note: This test controls the Type I experimentwise error for comparisons of all treatments against a control.

Alpha	0.05
Error Degrees of Freedom	175
Error Mean Square	0.032113
Critical Value of Dunnett's t	2.18341

Comparisons significant at the 0.05 level are indicated by ***.				
agegroup Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
2 - 1	0.02824	-Infinity	0.11924	
3 - 1	-0.01075	-Infinity	0.09564	
4 - 1	-0.11580	-Infinity	-0.02869	***
5 - 1	-0.25728	-Infinity	-0.16968	***

The multiple comparison results indicates that there are significant differences between the youngest group and group 4 (age 56-70) and between the youngest group and group 5 (age 71 or older).

Q.4: Use orthogonal polynomial contrasts to test for a linear, quadratic and cubic trend. What do you conclude?

```
contrast 'linear' agegroup -2 -1 0 1 2;
contrast 'quadratic' agegroup 2 -1 -2 -1 2;
contrast 'cubic' agegroup -1 2 0 -2 1;
run;
```

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
linear	1	1.72441448	1.72441448	53.70	<.0001
quadratic	1	0.37198661	0.37198661	11.58	0.0008
cubic	1	0.00373488	0.00373488	0.12	0.7335

The results for the linear, quadratic and cubic trend testing by orthogonal polynomial contrasts are shown above. The p-values for “linear” and “quadratic” are smaller than the significance level are smaller than 0.05, which means that the model has a linear and a quadratic trend, or there is a linear and a quadratic trend in the group means. Since the p-value for “cubic” is greater than 0.05, we can conclude that there is no cubic trend in this model (in the group means).

Q.5: Remove the last (oldest) group so that there are now 4 age groups. Determine the appropriate orthogonal polynomial contrasts to test for a linear and quadratic trend with 4 groups. Test the contrasts. What do you conclude?

```
proc glm data=upsit5;
	class agegroup;
	model smell = agegroup;
contrast 'linear' agegroup -3 -1 1 3;
contrast 'quadratic' agegroup 1 -1 -1 1;
run;
```

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
linear	1	0.28627815	0.28627815	12.22	0.0006
quadratic	1	0.14217420	0.14217420	6.07	0.0150

After removing the oldest group, do the orthogonal polynomial contrasts again to test if this model has linear or quadratic trend. The p-values for “linear” and “quadratic” are smaller than the significant level 0.05, we can conclude that this analysis shows highly significant linear and quadratic effects in the group means.

Biostat 200B HW7

2. $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, $i=1,\dots,a$, $j=1,\dots,n_i$

$Y_{ij} = \mu_i + \varepsilon_{ij} \Rightarrow \mu_i = \mu + \tau_i$

$\Rightarrow \mu = \mu_i - \tau_i$ (both take sum $\sum_{i=1}^{a}$ and multiply by $\frac{1}{a}$)

$\Rightarrow \frac{1}{a}\sum_{i=1}^{a}\mu = \frac{1}{a}\sum_{i=1}^{a}\mu_i - \frac{1}{a}\sum_{i=1}^{a}\tau_i$

$\Rightarrow \frac{1}{a}\cdot a\mu = \frac{1}{a}\sum_{i=1}^{a}\mu_i - \frac{1}{a}\sum_{i=1}^{a}\tau_i$ (constraint : $\sum_{i=1}^{a}\tau_i = 0$)

$\Rightarrow \mu = \frac{1}{a}\sum_{i=1}^{a}\mu_i - \frac{1}{a}\cdot 0 = \frac{1}{a}\sum_{i=1}^{a}\mu_i$ #

μ is the mean of the cell means, also called the grand mean.

3. $\bar{x}_1 = 27$, $\bar{x}_2 = 26$, $\bar{x}_3 = 23$, $\bar{x}_4 = 30$, $S_p^2 = 14$ $n=4$, $N=16$, $m = \binom{4}{2} = \frac{\overset{2}{4}\times 3}{2\times 1} = 6$ (pairwise)

(1) Bonferroni method (let α be 0.05)

The Bonferroni significant difference $BSD = t_{1-\frac{0.05}{12}, 12}\sqrt{14\cdot\frac{2}{4}}$

$= t_{.99583, 12}\cdot\sqrt{7}$

$= 3.152251\cdot\sqrt{7} \approx 8.34$

Means that are greater than 8.34 units apart are significantly different.

	$\bar{x}_3$	$\bar{x}_2$	$\bar{x}_1$	$\bar{x}_4$
mean	23	26	27	30

$\Rightarrow$ The means of each of those 4 groups are not statistically different.

(2) Tukey-Kramer method (let α be 0.05)

Tukey's honest significant difference $HSD = q_{1-0.05, 4, 12}\sqrt{14\left(\frac{1}{4}\right)}$

$= 4.199\cdot\sqrt{\frac{7}{2}} \approx 7.855$

Differences between group means greater than 7.855 units will be deemed significantly different.

	$\bar{x}_3$	$\bar{x}_2$	$\bar{x}_1$	$\bar{x}_4$
mean	23	26	27	30

$\Rightarrow$ The means of each of those 4 groups are not statistically different.