1. Diet plans – no SAS

(a)

Source	d.f.	Sum of Squares	Mean Square	F
Treatment	3	2272.80	757.60	34.33
Error	20	441.35	22.07	
Total	23	2714.15		

Note: remember that the total SS = treatment SS + error SS and total d.f. = treatment d.f. + error d.f.

(b) 4 treatments.

Note: treatment d.f. = r – 1.

(c) There are 24 observations in total. If all groups have the sample size, there are 6 observations per treatment.

Note: this only works if n's are the same for all groups. If the n's were unequal, the number observations in each group can not be determined from the ANOVA table.

(d) $F = 34.33 > F_{3,20,0.999} = 8.10$. Therefore, we reject the null hypothesis that all means are equal. The p-value is < 0.001. There is very strong evidence that at least one mean is different.

2. Fixed or random effects.

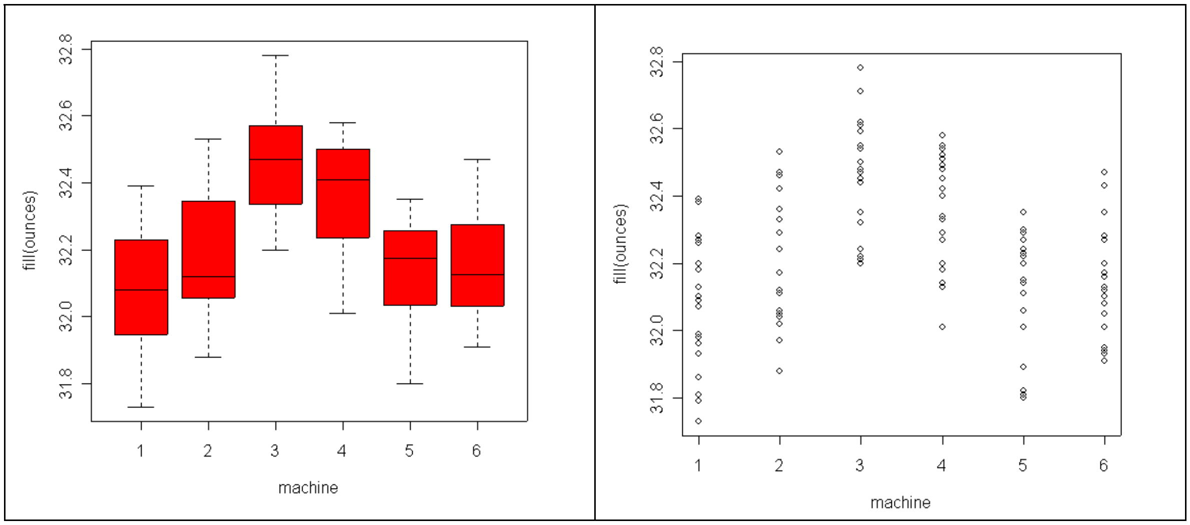
(a) Schools are random. There is no interest in a specific school.
(b) Schools are fixed. You want to compare school A to each of the other 9 schools.
(c) Schools are random. There is no interest in specific schools, but the variability among schools within a curriculum will be an important part of the analysis. Curriculum is fixed. You want to estimate differences between specific curriculums.

3. Bottle filling. My SAS code, without extra run commands:

```
data bottle;
  infile 'c:/philip/stat500/data/bottle.txt';
  input fill machine;    /* can ignore other two variables */
proc boxplot;
  plot fill*machine;
proc gplot;     /* hi resolution dot plots */
  plot fill*machine;
proc glm;
  class machine;
  model fill = machine ;
  lsmeans machine /stderr pdiff cl;
```

```
  estimate 'Contrast a' machine 0.25 0.25 -0.5 -0.5 0.25 0.25;
  estimate 'Contrast b' machine 0.5 0.5 0 0 -0.5 -0.5;
  contrast 'a and b' machine 1 1 -2 -2 1 1,
                     machine 1 1 0 0 -1 -1;
  contrast 'everything else'
   machine 1 -1 0 0 0 0,
   machine 0 0 1 -1 0 0,
   machine 0 0 0 0 1 -1;

  /* contrasts for 'everything else' are not necessary */
  /*  I expected you to get the SS that you need by difference */

  output out=resids r=resid p=yhat;
  title 'Bottle filling problem, hw 4';
  run;

/* I anticipated you would calculate parts a and b by hand */
/* perhaps useful bit of SAS trivia: */
/*  /clparm on model statement, e.g. */
/*     model fill = machine/clparm ; */
/* provides confidence intervals (default 95%) */
/*  for all 'Parameters', i.e. estimates and regression slopes */

proc plot;
  plot resid*yhat;
  run;

proc univariate normal;
  var resid;
  qqplot;         /* gives a hi-res qqplot */
  run;
```

(a)



Both the boxplot or the dot plot (above) show the location and spread of the data for each machine. It isn't reasonable that all machines have the same mean.

(b)

The GLM Procedure

Dependent Variable: fill

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	2.28934667	0.45786933	14.78	<.0001
Error	114	3.53060000	0.03097018		
Corrected Total	119	5.81994667			

F = 14.78, p < 0.0001. There is strong evidence of that at least one machine has a different mean.

(c) LSMEANS output gives you what you need for the test

machine	LSMEAN	Standard Error	Pr > |t|
1	32.0735000	0.0393511	<.0001
2	32.1905000	0.0393511	<.0001
3	32.4600000	0.0393511	<.0001
4	32.3655000	0.0393511	<.0001
5	32.1250000	0.0393511	<.0001
6	32.1515000	0.0393511	<.0001

Using a one sample t-test. $T = \frac{32.0735 - 32.1}{\sqrt{\frac{MSE}{n_1}}} = \frac{32.0735 - 32.1}{\sqrt{\frac{0.031}{20}}} = -0.673$, which is less extreme than the 0.975 t-quantile for a 114 d.f. T distribution = 1.981. Hence, there is no evidence that the mean fill for machine 1 differs from 32.1 oz.

(d) obtained from SAS output using lsmeans machine / pdiff cl; or by hand:

$$(\mu_5 - \mu_6) \pm t_{114,0.975}\sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_1}\right)} = (32.125\text{-}32.1515) \pm 1.981\sqrt{0.031 \times \frac{2}{20}}$$

$= \left(\text{-}0.137, 0.084\right)$, which is the 95% confidence interval for the mean difference between machines 5 and 6.

(3) SAS code provided in assignment

For Model I:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	124634.6999	124634.6999	2548396	<.0001
Error	119	5.8199	0.0489		
Uncorrected Total	120	124640.5198			

For Model II:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	2.28934667	0.45786933	14.78	<.0001
Error	114	3.53060000	0.03097018		
Corrected Total	119	5.81994667			

From the output in SAS, $SS_1 = 5.8199$, $SS_2 = 3.5306$, and $df_1 = 119$, $df_2 = 114$, then the F statistic comparing these two models should be

$$F = \frac{(5.8199 - 3.5306)/(119 - 114)}{3.5306/114} = 14.7839 > 4.43289 = F_{5,114,0.999}.$$

This is the same F statistic you got in part b.

(f) The SAS output for the estimate statement provides the information:

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Contrast a	-0.27762500	0.03407905	-8.15	<.0001
Contrast b	-0.00625000	0.03935110	-0.16	0.8741

* To report this on an exam or in a scientific report, I would round the numbers to the 0.01 digit. This is because the s.e/3 = 0.01.

(g) Yes, the contrasts are uncorrelated (or orthogonal). The sample sizes are equal, so the short form of the orthogonality criterion is sufficient. The sum of cross-products of contrast coefficients = 0.

(h) The SS for the 'interesting' parts from the contrast statement labeled 'a and b' above is 2.056, so the left over SS = 2.289 – 2.056 = 0.233. This has df = 5 – 2 = 3. Hence the F statistic for the 'left over SS' = (0.233 / 3) / MSE = 2.51. Under H0: no leftover effects, this has an F 3, 114 distribution. Using tables, p > 0.05. Using the computer p = 0.062.

4: Diet and longevity My SAS code for all numerical parts:

```
data diet;
  infile 'c:/philip/stat 500/data/dietlong.txt';
  input longevity diet $;

proc glm;
  class diet;
  model longevity = diet;
  lsmeans diet /pdiff adjust=tukey;

  estimate '1' treatment 0   -0.25 -0.25 1   -0.25 -0.25;
  estimate '2' treatment 0   0      1    0   -1     0;
  estimate '3' treatment 0   0      1    0    0    -1;
  estimate '4' treatment 1   0      -0.5 0    0    -0.5;
  estimate '5' treatment -1  0      0    1    0     0;

  run;
```

(a) This question calls for the 'omnibus' F test.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	12733.94181	2546.78836	57.10	<.0001
Error	343	15297.41532	44.59888		
Corrected Total	348	28031.35713			

The p-value of the F-test is <.0001. There is very strong evidence that at least one mean lifetime (in months) is different.

(b) The specified family is all pairwise comparisons, so the most appropriate procedure is Tukey (or Tukey-Kramer). Using that method, there are lots of significant differences at an experiment-wise 5% error rate:
N/N85 is significantly different from every other treatment.
N/R40 is significantly different from NP and lopro.
N/R50 is significantly different from NP
NP is significantly different from every other treatment.

(c)

Contrast	Interpretation
1	The average effect of caloric limitation, relative to the unrestricted diet.
2	The effect of reduced calories before weaning. (could add: in diets with restricted post-weaning calories)
3	The effect of low protein content post-weaning (in diets with restricted post-weaning calories)
4	The effect of reducing from 85 kcal/wk to 50 kcal/wk post-weaning, assuming that protein restriction has no effect.
5	The effect of ad lib food relative to a normal diet.

Other wordings are very possible. My goal in asking this part was to get you to think about interpretation of contrasts

(d) Yes, 1 and 2 are orthogonal. Yes, 3 and 4 are orthogonal

(e) No for both, (-0.25*1)/71 + (-0.25*-1)/56 = 0.000943, and (-0.5*-1/71) + (-0.5*1/56) = -0.000943

(f)

Parameter	Estimate	Error	t Value	Pr > |t|
1	-15.0942788	1.04663700	-14.42	<.0001
2	-0.5885312	1.19355007	-0.49	0.6223
3	2.6114688	1.19355007	2.19	0.0293
4	-8.3002206	1.06704119	-7.78	<.0001
5	-5.2891873	1.30100640	-4.07	<.0001

The biggest problem here is making sure that the coefficients go in the correct columns. The order of the diets (as given by the class statement output or the lsmeans output) is:

```
N/N85   N/R40   N/R50   NP   R/R50   lopro

If lopro were capitalized it would be first in the list.
```