1. Small experiment, hand calculations

   (a) F = 7.69, F distribution with 4,5 df.
   Using the SS provided:

Source		df	SS	MS	F
Trt	$\sum_i n_i(\overline{y}_i - \overline{y}_.)^2$	4	40.25	10.0625	7.69
Error	$\sum_{ij}(y_{ij} - \overline{y}_i)^2$	5	6.54	1.3080	
C.total	$\sum_{ij}(y_{ij} - \overline{y}_.)^2$	9	46.79		

   Note: I gave you all three SS. Any two could be used. For example, the trt SS could be computed from the full model SS (error above) and reduced model SS (c.total above).

   (b) -3, -2, -1, 1, 5
   The coefficients are $x_i - \overline{x}_.$ = -7.5, -5, -2.5, 2.5, 12.5. Dividing by 2.5 makes them nice. If you want to really want to estimate the regression slope, the coefficients are -7.5/250, -5/250, -2.5/250, 2.5/250 and 12.5/250. Any of these three sets accepted for full credit. Your answers to part 1d will differ, but the SS and F in parts 1e and 1f will be the same. Note: Yes, these are the same as the backup set.

   (c) Yes, because $\sum_i l_i k_i = 0$: $0 \times -3 + 2 \times -2 + -3 \times -1 + 1 \times 1 + 5 \times 0 = 0$

   (d) $\hat{\gamma} = 26.27$, se = 5.11
   $\hat{\gamma} = -3 \times 3.48 + -2 \times 5.19 + -1 \times 6.02 + 1 \times 5.31 + 5 \times 9.56 = 26.27$
   Var $\hat{\gamma} = 1.308 \sum l_i^2/2 = 1.308 \times 40/2 = 26.16$

   (e) 34.50
   $\text{SS} = \hat{\gamma}^2 / \sum l_i^2/n_i = 26.27^2/20 = 34.50$

   (f) F = 1.46, F distribution with 3,5 df.

Source	df	SS	MS	F
Trt	4	40.25		
Contrast	1	34.50		
Left-over	3	5.74	1.91	1.46
Error	5		1.308	

   (g) F = 13.49, F distribution with 2,5 df. There are two ways to compute this:

      i. A $\boldsymbol{C\beta}$ test. $\boldsymbol{C\beta} = \begin{bmatrix} 26.27 \\ -2.37 \end{bmatrix}$ using a cell means model, $\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$, so $\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$ and $(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \end{bmatrix}$

$\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}' = \begin{bmatrix} 20 & 0 \\ 0 & 7 \end{bmatrix}$, and $\left[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\right]^{-1} = \begin{bmatrix} 1/20 & 0 \\ 0 & 1/7 \end{bmatrix}$,

so $(\boldsymbol{C\beta} - \boldsymbol{m})' \left[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\right]^{-1} (\boldsymbol{C\beta} - \boldsymbol{m}) = (26.27)^2/20 + (-2.37)^2/7 = 35.31$, and

$$F = \frac{35.31}{2 \times 1.308} = 13.49.$$

ii. Or, by recognizing that the two components are orthogonal, so you can compute SS for each and add. You already have SS for the first component: 34.50. That for the second is $(-2.37)^2/7 = 0.802$. The numerator is $34.50 + 0.802 = 35.31$. So,

$$F = \frac{35.31}{2 \times 1.308} = 13.49.$$

Note: I revised this question after writing the first version of the exam. The first version specifically asked for a $\boldsymbol{C\beta}$ test. When I marked answers, I initially forgot that revision. I believe I caught all of those grading mistakes. If I took off 5 points for "not doing a $\boldsymbol{C\beta}$ test, I missed you. See me and I will regrade your answer.

2. The moon and mental health

(a) 1) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}$, 2) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}$.
Note: 2) is exactly the same as 1) because I only asked for a small piece of the $\boldsymbol{X}$ matrix. For this piece, the column for Apr contains 3 ones.

(b) No - the type I and type III SS are not the same when you look at the model with interaction or both SS in the additive model. Noting that there are 42 observations, which doesn't evenly divide into 36 groups was also accepted for full credit.

(c) No - the interaction df is the product of the two main effect df.

(d) F = 9.21, central F with 2,6 df.
This is the type III SS for phase in the model with month, phase, and interaction. I believe the model with month*phase interaction is the most appropriate. That is equivalent to the cell means model, followed by specific contrasts. The F calculated from sequential SS (9.01) does not correspond to contrasts following the cell means model because the data are unbalanced. The F from the additive model (phase and month only) = 6.06 was accepted for almost full credit. The F from the phase only model (F = 1.61) got no credit because the problem says "Admissions are known to vary monthly for many reasons …". A reasonable model needs to account for variation among months, i.e. by blocking on month by including month in the model.

(e) Removed from exam. I did this because the se computation can not be done without further information. The data are unbalanced. We haven't talked about how to compute the se in this case. The se depends on the number of months with 1 obs per phase, the number of months with 1 obs for one phase and 2 for the other, and (potentially) the number with 2 obs for each phase. Even with that information, the se computation is laborious.

(f) # months = 60
The key insights are that the # observations is twice the number of months and k for the sample size calculation is $\sqrt{2}$ because the se of the difference of two means is $\sigma\frac{\sqrt{2}}{n}$.

$$\begin{aligned} n &= \#obs = (1.973 + 0.844)^2(\sqrt{2})^2\frac{4.2}{0.75^2} \\ &= 118.4 \end{aligned}$$

So # months = 59.2, i.e. 60.
A couple of folks got to 60 by compensating errors: omitting k and forgetting to convert from observations to months. That got docked points for each mistake even though the answer turned out to be correct. The process was more importnant than the final answer.

(g) error df = 180
The new study would have 60 months, 3 phases, and 2 obs per month/phase, for a total of 360 observations. The model will have 59 + 2 + 118 df, leaving 359-179 = 180 df.
If you used the backup value of 40 months, you should have gotten 239 - 119 = 120 df for error.
Quite a few people correctly determined the new number of observations in the previous part but forgot that there were more than 12 months in the new study. If you used month df = 11 and interaction df = 22 in the error df computation, you made this mistake.

3. Likelihood and home sales

(a) smaller in A
The variance is the inverse of the negative 2nd derivative. A is more sharply curved (larger 2nd derivative) at the mle.

(b) find the value of the log likelihood at the mle of β_1
subtract one-half of the 0.99 quantile of a χ^2_1 distribution
find what parameter values for β_1 result in that log likelihood
Note: arguing that the curve is symmetrical, so the Wald and profile intervals are similar got partial credit because "similar" is not "the same" unless the lnL trace is perfectly quadratic (which I can't tell from the figure).

(c) quite different because the profile lnL trace is not symmetric

(d) Z = -4.19, standard normal

(e) odds ratio = 2.74
$\hat{\beta}_1 = -0.101$, that is the change in log odds per increase of 1 $ per sq. foot. The change for a decrease of 10 $ per sq. foot is (-10)(-0.101) = 1.01. The odds ratio is $\exp 1.01 = 2.74$.

(f) (1.76, 4.26)
The 95% ci for the log odds for a decrease of 10 $ per sq. foot is $1.01 \pm 1.96 * (-10) * 0.024 = (0.56,\ 1.45)$. The 95% ci for the odds ratio is $(\exp 0.56,\ \exp 1.45) = (1.76,\ 4.26)$.

(g) P[sell in 3 months] = 0.98
At the stated conditions, $\boldsymbol{X\beta} = 13.35 + 90 * (-0.1007) + 10 * (-0.0519) = 3.768$. $P = \frac{1}{1+\exp(-\boldsymbol{X\beta})} = \frac{1}{1+\exp(-3.768)} = 0.98$.