1) A one-factor repeated-measures ANOVA (with three levels of time) on errors revealed a significant difference across time points, $F(2, 8) = 12.16, p = .015$ (Huynh-Feldt adjusted), $\eta^2_p = .75$, with errors highest at time 1 ($M = 10.2$) and decreasing at time 2 ($M = 8.0$) and time 3 ($M = 6.2$).

b) A contrast of time 1 versus times 2 and 3 combined showed that errors were significantly higher at time 1 ($M = 10.2$) than they were later ($M = 7.1$), $t(4) = 3.3, p = .03$.

c) The analysis in part b is more specific, allowing a more precise conclusion. Also, there are no concerns about sphericity when performing single-$df$ contrasts.

2) A 3 (stimulus valence: positive, negative, neutral) x 2 (beverage: wine, water) repeated-measures ANOVA was conducted on ratings. Ratings for positive stimuli ($M = 25.4$) were highest, followed by those for neutral stimuli ($M = 12.7$), which was followed by those for negative stimuli ($M = -12.0$); this main effect was significant, $F(2, 38) = 206.7, p < .001, \eta^2_p = .92$. There was also a significant preference for the wine advertisements ($M = 8.7$) over the water advertisements ($M = 3.8$), $F(1, 19) = 28.5, p < .001, \eta^2_p = .60$. These main effects were qualified by significant interaction, $F(2, 38) = 81.5, p < .001, \eta^2_p = .81$; wine ratings were higher than those for water for positive and neutral stimuli, but were lower for negatively-valenced stimuli (see the figure at right).

b) Testing the simple effects of beverage at each level of stimulus will allow more-detailed claims to be made.

c) Three simple-effect tests were executed using paired-sample t-tests (using a Bonferroni-corrected alpha level of $.05 / 3 = .017$). For positive stimuli, wine ($M = 25.4$) was rated significantly higher than water ($M = 10.2$), $t(19) = 8.97, p < .001, d = 2.0$; the same was true for neutral stimuli (wine $M = 12.7$, water $M = 2.7$), $t(19) = 8.00, p < .001, d = 1.8$. However, there was a significant difference in the opposite direction (wine $M = -12.0$, water $M = -1.4$) for negative stimuli, $t(19) = 6.58, p < .001, d = 1.5$.

3) A 3 (maze: radial, water, T-maze) x 2 (surgery: critical, control) mixed-factor ANOVA (with the first factor being a repeated-measures factor) was conducted on maze scores (with lower scores being assumed to indicate worse performance). The critical area group ($M = 9.8$) performed non-significantly worse than the control group ($M = 11.4$), $F(1, 6) = 0.87, p = .39$. Even using the Huynh-Feldt correction for nonsphericity, there was a significant effect of maze, $F(2, 12) = 5.27, p = .02, \eta^2_p = .47$, with performance in the water ($M = 11.5$) and T-maze ($M = 11.3$) being higher than in the radial maze ($M = 9.1$). This effect is qualified by a significant maze-by-surgery condition interaction, $F(2, 12) = 6.02, p = .02, \eta^2_p = .50$, which appears to be due to damage to the critical region specifically impairing performance on the radial task, but not apparently the other two maze tasks.

c) The error term for the within-subjects effects is SM/R. (Sorry to have used A and B then R and M.)

d) The error term for the between-subjects effects is S/R.

e) The main effect of maze-type might be usefully explored with some pairwise comparisons.

f) Bonferroni-corrected pairwise comparisons revealed that – despite the significant effect of maze – there were no significant pairwise comparisons ($p \geq .054$); this can happen. A non-pairwise comparison of the radial maze against the other two produces $p = .006$; syntax for this appears at right above.

g) Some simple-effect tests of the different surgery conditions (within each maze) would be useful.

h) Three simple-effect tests were executed using independent-sample t-tests (with a Bonferroni-corrected alpha level of $.05 / 3 = .017$). Only the group difference in the radial maze (control $M = 11.5$, critical-area $M = 6.8$) approached significance, $t(6) = 3.02, p = .02$. In the other mazes the group differences were not significant ($p > .8$). (This is one of those cases where using an error term based on the omnibus analysis would offer a little more power, which might be just enough to nudge the radial-maze effect into significant territory; see your textbook for more about this.)

GLM radial water t_maze BY cond /
WSFACTOR=maze 3 /
/MMATRIX radial 1 water -1/2 t_maze -1/2 /
DESIGN=cond.
4) A correlation analysis revealed a strong, positive relationship between SAT-like guessing scores and actual SAT verbal scores, \( r(26) = .53, \) \( p = .004. \) This suggests that a fair amount (28\%) of the variance in SAT verbal scores can be predicted by whatever test-taking (or other) strategies students are using to guess. To the extent that these strategies have nothing to do with reading comprehension, per se, this casts at least some doubt on the construct validity of SAT verbal scores, no? (These data are realistic and are based on a real experiment.)

b) In this case we are testing the null hypothesis \( H_0: \rho_1 = \rho_2. \) The appropriate equation to use \((Z_{r1} = .59, Z_{r2} = .829)\) is:

\[
Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} = \frac{.59 - .829}{\sqrt{(1/25) + (1/14)}} = -.72
\]

This is much less than the critical value of ±1.96, so we must conclude that there is no significant difference in the SAT-like/SATV relationship whether or not people have read the passages.

c) \( R^2 = .283. \) This is the amount of variance in SAT scores that is predictable knowing the SATlike-SATV relationship.

d) \( s_{Y,X} = 53.1. \) This is the average (typical, standard) error of prediction when using SATlike scores to predict SATV scores.

5) a) \( df_{total} \) should be 89 (total observations minus 1); if you used \( n \) instead of \( i \) to represent the number of subjects, no worries

<table>
<thead>
<tr>
<th>SV</th>
<th>df</th>
<th>EMS</th>
<th>( F ) (error term)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Between subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>( a - 1 = 2 )</td>
<td>( e + A + I/AS )</td>
<td>( I/AS )</td>
</tr>
<tr>
<td>S</td>
<td>( s - 1 = 1 )</td>
<td>( e + G + I/AS )</td>
<td>( I/AS )</td>
</tr>
<tr>
<td>AS</td>
<td>( a - 1)(s - 1) = 2 )</td>
<td>( e + AG + I/AS )</td>
<td>( I/AS )</td>
</tr>
<tr>
<td>I/AS</td>
<td>( as(i - 1) = 24 )</td>
<td>( e + I/AS )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Within subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>( d - 1 = 2 )</td>
<td>( e + D + DI/AS )</td>
<td>( DI/AS )</td>
</tr>
<tr>
<td>DA</td>
<td>( d - 1)(a - 1) = 4 )</td>
<td>( e + DA + DI/AS )</td>
<td>( DI/AS )</td>
</tr>
<tr>
<td>DS</td>
<td>( d - 1)(g - 1) = 2 )</td>
<td>( e + DG + DI/AS )</td>
<td>( DI/AS )</td>
</tr>
<tr>
<td>DAS</td>
<td>( d - 1)(a - 1)(s - 1) = 4 )</td>
<td>( e + DAG + DI/AS )</td>
<td>( DI/AS )</td>
</tr>
<tr>
<td>DI/AS</td>
<td>( as(d - 1)(i - 1) = 48 )</td>
<td>( e + DI/AS )</td>
<td></td>
</tr>
</tbody>
</table>

b. i) A
b. ii) D/A2

c. i) Does creativity change across delays?
c. ii) Does the effect of amount of alcohol consumed differ for men and women?
6) a) For the full sample, \( r = .32 \).
   b) When \( \text{grouping}1 = 0 \), \( r = .19 \).
   c) When \( \text{grouping}1 = 1 \), \( r = .09 \).
   d) The correlations in b and c are both smaller than in a.
   e) When \( \text{grouping}2 = 0 \), \( r = .50 \).
   f) When \( \text{grouping}2 = 1 \), \( r = .39 \).
   g) The correlations in e and f are both larger than in a.
   h) The \( \text{grouping}1 \) variable leads to a restriction of range issue. The SD of \( x \) for the full sample is 8.6, but within \( \text{grouping}1 = 0 \) it’s 4.2 and within \( \text{grouping}1 = 1 \) it’s 6.4.

   The \( \text{grouping}2 \) variable leads to there being less variability around the regression line within the subgroups than within the full group. Less \( y \) variability means (in this case at least, where \( x \) variability is similar in the groups created by \( \text{grouping}2 \)) better fit and a stronger linear relationship. The scatterplot at right may help illustrate this. The red line is the regression line for the full sample, and the black and gray lines are the regression lines for the subsamples (which are illustrated with black and gray). Notice that slope is relatively stable in all three groups.

7) a) \( R^2 = .283 \); \( SS_{\text{regression}} = 1552.957 \)
   b) \( R^2 = .283 \); \( SS_{\text{regression}} = 128648.035 \)
   c) region b
   d) region b
   e) \( R^2 \) is the proportion of \( SS_{\text{total}} \) (for the DV) that is associated with the predictor (i.e., \( SS_{\text{regression}} \)). The same \( R^2 \) results from different values of \( SS_{\text{regression}} \) because the denominator (and the numerator, for that matter) changes with the units of measurement in the DV. SSs are not quantified in scale-free units; they are based on the (squared) units of measurement of the DV (and are positively correlated with sample size), meaning that a single SS is uninterpretable as being small, large, or otherwise. (Dividing an SS turns into something more interpretable: variance. Finding the square root of a variance is perhaps even a more interpretable thing: \text{SD}.)