1. An experiment was conducted to ascertain gas mileage obtained with three grades of gasoline: regular, extra and premium. Four cars were used in the experiment, with all three gasolines being used (at different times) in each of the four cars. The order in which a given car received the three gasolines was determined randomly. The response variable was miles per gallon of gasoline. The following (partially completed) ANOVA table was obtained from this experiment's data.

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Gasolines		47.17		
Cars	3	390.25		
Error				
Total	11	448.92		

Use this information to answer the following **3 questions** (the current one and the next 2).

The type of experimental design used here was:

(a) A completely randomized design.

(b) A randomized block design with gasolines as blocks.

(c) A randomized block design with cars as blocks.

(d) A block design but not randomized.

(e) A crossover design.

This is clearly a block design with the cars being the 4 blocks each of which are assigned to all 3 of the gasoline grades (the treatments) at different times, which is the very essence of a block design (see pg. 162-163 of Chapter 2B notes), and for each car, the order in which the 3 treatments were assigned was determined randomly. So this is a randomized block design (RBD). This example is very similar to the illustration of RBDs given in class (pg. 167-170).

2. The value of the F-statistic and an accurate range of the P-value of the F-test for testing whether mean miles per gallon differ significantly between gasolines are given by:

(a) F-statistic $= 8.2$; and P-value < 0.05.

(b) F-statistic $= 8.2$; and $0.05 < P$-value < 0.1.

(c) F-statistic $= 12.3$; and P-value < 0.05.

(d) F-statistic $= 12.3$; and $0.05 < P$-value < 0.1.

(e) F-statistic $= 10.1$; and P-value < 0.05.

This is an RBD setting with $k = 3$ treatments (gasolines) and $n = 4$ blocks (cars). Using pg. 166-167 of Chapter 2B notes and the partial ANOVA table given, the SSE is given by:

$$SSE = SST - SSTr - SSB = 448.92 - 390.25 - 47.17 = 11.5,$$

The degrees of freedom for the treatments and error are given by $k-1 = 2$ and $(k-1)(n-1) = 6$ respectively. Hence, the F-statistic for testing equality of the treatment means equals:

$$F \equiv F_{Tr} = \frac{MSTr}{MSE} = \frac{SSTr/(k-1)}{SSE/\{(k-1)(n-1)\}} = \frac{47.17/2}{11.5/6} = 12.3.$$

Under H_0, F_{Tr} follows a $F_{k-1,(k-1)(n-1)} \equiv F_{2,6}$ distribution and we reject H_0 for large values of F_{Tr}. Hence, the P-value is $P(F_{2,6} > F_{obs}) = P(F_{2,6} > 12.3)$, where $F_{obs} = 12.3$ denotes the observed value of the test statistic F_{Tr}. From the F-table, we have: $F_{2,6;0.05} = 5.14 < 12.3$, so that $P(F_{2,6} > 12.3) < P(F_{2,6} > 5.14) = 0.05$. Hence, the P-value is less than 0.05.

3. The mean miles per gallon for regular, extra and premium were 24.3, 27 and 30.25, respectively. Using Tukey's procedure with an experimentwise error rate (EWER) of 0.05, which of the following is the correct conclusion?

(a) Each gasoline is significantly different from the other two with respect to mean miles per gallon.

(b) Premium yields significantly higher mean miles per gallon than both regular and extra, but regular and extra do not differ significantly from each other.

(c) We cannot conclude that there are any differences between the gasoline means.

(d) There is an interaction between cars and gasolines.

(e) Premium yields significantly higher mean miles per gallon than regular, but neither premium and extra, nor regular and extra, differ significantly from each other.

The difference between any two gasoline grade means should be compared with the appropriate LSD value obtained from applying Tukey's HSD procedure for pairwise comparison of the gasoline grade means. Using pg. 169 of Chapter 2B notes, this LSD value is given by:

$$Q_{\alpha,k,(k-1)(n-1)}\sqrt{\frac{MSE}{n}} = Q_{0.05,3,6}\sqrt{\frac{SSE/6}{4}} = 4.34\sqrt{\frac{11.5/6}{4}} = 4.34\sqrt{\frac{1.917}{4}} = 3.00.$$

Based on this LSD value, the mean miles per gallon for premium is thus significantly more than that for either regular or extra, but regular and extra do not differ significantly. The absolute differences in means for (premium, regular), (premium, extra) and (regular, extra) are 5.95, 3.25 and 2.7, respectively. Only the first two exceed the LSD value obtained above.

4. We discussed the use of a log-transformation of the response variable in a regression analysis. Doing so can be beneficial because:

(a) It can sometimes induce a linear model for the mean.

(b) It can sometimes make the variance of the response variable constant.

(c) Both (a) and (b) are true.

(d) The resulting parameter estimates are always unbiased.

(e) It pleases lumberjacks.

See pg. 74-76 of Chapter 1B notes. We discussed the potential scenarios and the benefits of applying log transformations in details several times in class. Since this was straight out of the notes, you will **not** get any partial credit for picking out only one of the answers, i.e. (a) or (b). You get credit only if you mark the *most* correct and inclusive option which is (c).

5. Over the past five years an insurance company has had a mix of 40% Whole Life policies, 20% Universal Life policies, 25% Annual Renewable-Term (ART) policies, and 15% other types of policies. A change in this mix over the long haul could require a change in the commission structure, reserves, and possibly investments. A sample of 1000 policies issued over the last few months gave the following results.

Category	Number of Policies
Whole Life	320
Universal Life	280
ART	240
Other	160

Use this information to answer the following **2 questions** (the current and the next one).

It turns out that at a level of significance 0.05, one can reject the hypothesis that the policies issued in the last few months match the historical percentages. Which of the following is true of the χ^2 test used to reach this conclusion?

(a) The test statistic is 10.71 and the P-value is between 0.01 and 0.025.

(b) The test statistic is 43.90 but the χ^2 table tells us nothing about the P-value.

(c) The test statistic is 43.90 and the P-value is less than 0.005.

(d) The test statistic is 49.07 but the χ^2 table tells us nothing about the P-value.

(e) The test statistic is 49.07 and the P-value is less than 0.005.

This corresponds to a Multinomial experiment with $k = 4$ categories and the χ^2 goodness-of-fit test with category probabilities fully specified. Under H_0, i.e. assuming no change in the historical percentages, the expected numbers of policies in the four categories are as follows:

$$0.40(1000) = 400, \quad 0.20(1000) = 200, \quad 0.25(1000) = 250 \quad \text{and} \quad 0.15(1000) = 150.$$

Using pg. 179-181 of Chapter 3 notes, the χ^2 statistic is thus

$$\chi^2 = \frac{(320-400)^2}{400} + \frac{(280-200)^2}{200} + \frac{(240-250)^2}{250} + \frac{(160-150)^2}{150} = 49.07.$$

Under H_0, $\chi^2 \sim \chi^2_{k-1} \equiv \chi^2_3$, and we reject H_0 for large values of χ^2. Thus, the P-value is $P(\chi^2_3 > \chi^2_{obs})$ $= P(\chi^2_3 > 49.07)$, where $\chi^2_{obs} = 49.07$ is the observed value of the test statistic. Comparing this value with the percentiles of the χ^2_3 distribution, we observe that $\chi^2_{3;0.005} = 12.838$ and $49.07 > 12.838$, and so, P-value $= P(\chi^2_3 > 49.07) < P(\chi^2_3 > 12.838) = 0.005$.

6. Which of the following is the best conclusion as to why the χ^2 test was significant?

(a) None of the percentages has changed very much.

(b) All four percentages have changed substantially.

(c) The percentages of Whole Life and Universal Life policies have both increased substantially.

(d) The percentages of ART and other policies haven't changed much, but the percentages of Whole Life and Universal Life have decreased and increased respectively.

(e) The insurance salesman who collected the data fudged the numbers.

The current percentages are 32, 28, 24 and 16, and so it is evident that the first two are the most different from the historical percentages.

7. In a regression setting, when the variance of a positive response variable Y tends to increase as x increases, a good way to try to rectify this problem is to:

(a) Multiply each Y-value by the same large, positive number.

(b) Divide each Y-value by the same large, positive number.

(c) Use $1/x$ as the independent variable.

(d) Take the natural logarithm of each Y-value.

(e) Use both x and x^2 in the regression model.

See pg. 74-76 of Chapter 1B notes. We discussed the potential scenarios and the benefits of applying log transformations in details several times in class.

8. An engineering group is interested in the effects of type of paint primer (A, B or C) and application method (dipping or spraying) on paint adhesion. Which of the following necessarily describes an interaction between primer type and application method?

(a) Primer B results in better adhesion than either primer A or C.

(b) Spraying results in better adhesion than dipping.

(c) Application method has no effect on adhesion properties, but type of primer does have an effect.

(d) When dipping is used, primer has no effect on adhesion, but when spraying is used, primer A leads to better adhesion than do primers B and C.

(e) The fumes from primer A are more toxic than those from primers B and C.

See pg. 142 of Chapter 2B notes for a correct interpretation of interaction. We also discussed this in detail in class and went over examples when interaction is *not* present (e.g., pg. 145).

9. Consider the following model for 1-way ANOVA and answer the question asked thereafter:

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \ldots, k, \text{ and } j = 1, \ldots, n,$$

where the variance of each ϵ_{ij} is σ^2. Assuming all our usual conditions on this model, the standard error of $\bar{X}_1$ is given by:

(a) σ.

(b) $\sigma/\sqrt{n}$.

(c) σ^2/n.

(d) $\mu_1/\sqrt{n}$.

(e) σ^2.

When we have a random sample, we know that the standard error of the sample mean is the population standard deviation divided by the square root of the sample size. Here, all observations X_{ij} from group 1 (i.e. $i = 1$) have a common standard deviation σ and there are n such observations. Hence, the standard deviation of their sample mean $\bar{X}_1$ is $\sigma/\sqrt{n}$.

10. Four hypothesis tests are conducted independently of each other. If the level of significance of each test is 0.10, then the experimentwise error rate is:

(a) 0.10.

(b) 0.1921.

(c) 0.2916.

(d) 0.3439.

(e) 0.6561.

The experimentwise error rate (EWER) is the probability of rejecting one or more null hypotheses when in fact all the null hypotheses are true. Let R_i denote the event that the ith hypothesis is rejected, $i = 1, 2, 3, 4$. Then, using pg. 135-136 of Chapter 2A notes,

$$\begin{aligned} EWER = P(R_1 \cup R_2 \cup R_3 \cup R_4) &= 1 - P(R_1^c \cap R_2^c \cap R_3^c \cap R_4^c) \\ &= 1 - P(R_1^c)P(R_2^c)P(R_3^c)P(R_4^c) \\ &= 1 - (0.90)^4 \\ &= 0.3439. \end{aligned}$$

11. Regression data consisting of $n = 200$ cases were collected. The response variable is y and there were four independent variables: x_1, x_2, x_3 and x_4. The accompanying output (available as a 2-page pdf document attached with this exam) shows various information obtained from these data in an "R" session. Use the information and the attachment above to answer the following **6 questions** (i.e. the current one and the next 5 questions).

Let M_1 be the linear model containing only the variables x_1 and x_4, and M_2 the linear model containing all four independent variables. The value of the F-statistic for testing:

$$H_0 : \text{Correct model is } M_1 \quad \text{vs.} \quad H_a : \text{Correct model is } M_2$$

(a) Is 72.9.

(b) Is 97.4.

(c) Is 150.5.

(d) Is 294.8.

(e) Cannot be determined from the information given.

You should use the reduction method of testing here. From the "R" output, the relevant SSEs for the full and reduced models are $SSE_f = 83,792$ and $SSE_r = 146,443$, respectively. Hence, the required F-statistic is:

$$F = \frac{(SSE_r - SSE_f)/(4-2)}{SSE_f/(n-4-1)} = \frac{(146,443 - 83,792)/2}{83,792/195} = 72.9.$$

12. The proportion of variance in y explained by the model containing x_2 and x_3 is:

(a) 0.855.

(b) 0.940.

(c) 0.364.

(d) 0.636.

(e) 0.750.

This is simply the multiple R^2 given in the summary for the model with only x_2 and x_3.

13. Consider the following information obtained using "R":

Model number	Variables used	R^2
1	x_1	0.7452
2	x_2	0.0493
3	x_1, x_2	0.8548
4	x_1, x_2, x_3	0.8548
5	x_1, x_2, x_3, x_4	0.8567

Based on this information, the most parsimonious model that still provides a near optimum fit is given by model number:

(a) 1.

(b) 2.

(c) 3.

(d) 4.

(e) 5.

Model 3 has an R^2 that is barely smaller than those of models 4 and 5, while being substantially larger than those of 1 and 2. Therefore, model 3 fits the bill.

14. Consider the output for the model containing all four of the independent variables. Which of the following is the best conclusion to draw from the model utility test?

(a) All four of the independent variables are useful.

(b) The variable x_1 should definitely be included in the model.

(c) None of the independent variables is useful.

(d) At least one of the 4 independent variables is useful.

(e) The variable x_2 should definitely be included in the model.

The P-value for the model utility test is extraordinarily small, $2 \cdot 10^{-16}$, and therefore there is strong evidence that at least one of the four independent variables is useful. However, as is always the case with a model utility test, this is all that can be said from the result of the test (and nothing can be said about the importance of any specific predictor in the model).

15. The estimate of the error variance in the model containing x_1 and x_2 is:

(a) $\sum_{i=1}^{200}(y_i - \bar{y})^2/199$.

(b) 84,908.

(c) 0.8548.

(d) 20.76.

(e) 430.98.

From the given `R` output, the estimate of σ^2 is $SSE/(n-k-1) = 84,908/(200-2-1) = 431.0$.

16. Based on the model containing all four independent variables, which of the following is closest to a 95% confidence interval for β_2?

(a) $5.23 \pm 1.96(0.30)$.

(b) $-3.44 \pm 1.96(0.37)$.

(c) $-3.44 \pm 1.96(20.73)$.

(d) $-3.44 \pm 1.96(20.73/\sqrt{195})$.

(e) $-3.44 \pm 1.96(20.73)/\sqrt{\sum_{i=1}^{200}(x_{i2}-\bar{x}_2)^2}$.

The error degrees of freedom $(n-k-1)$ are large, 195, and so $t_{195,0.025} \approx z_{0.025} = 1.96$. From the summary in the `R` output, the estimate of β_2 is -3.44 and the estimated standard error of $\hat{\beta}_2$ is $\hat{\sigma}_{\hat{\beta}_2} = 0.37$. Hence, using the formula on pg. 94 of Chapter 1C notes, or in pg. 70 of the notes, the required 95% confidence interval for β_2 is: $\hat{\beta}_2 \pm t_{n-k-1,0.025}\hat{\sigma}_{\hat{\beta}_2} = -3.44 \pm 1.96(0.37)$.

17. A sports reporter has studied data on professional golfers and determined that the following model reasonably approximates the relationship between average driving distance (x, in yards) and average eighteen hole score (Y):

$$Y = 182.9 - 0.7394x + 0.001203x^2 + \epsilon,$$

where ϵ is Normally distributed with mean 0 and standard deviation 0.25. Suppose a pro golfer's average driving distance is 285 yards. What is the probability that his average eighteen hole score is more than 70.25?

(a) 0.500.

(b) 0.360.

(c) 0.640.

(d) 0.928.

(e) 0.072.

$Y|x$ has a Normal distribution with mean $182.9 - 0.7394x + 0.001203x^2$ and standard deviation 0.25 under the assumed model. Hence, given $x = 285$, and letting $Z \sim N(0,1)$, we have:

$$\begin{aligned} P(Y > 70.25) &= P\left(Z > \frac{70.25 - 182.9 + 0.7394(285) - 0.001203(285)^2}{0.25}\right) \\ &= P(Z > 1.463) \\ &= 0.072. \end{aligned}$$

18. Consider again the setting of Problem **17** involving golfers. What is the 10th percentile of the distribution of average scores for golfers whose average driving distance is 290 yards?

(a) 68.5.

(b) 69.3.

(c) 69.8.

(d) 70.6.

(e) 71.2.

Under the assumed model, the mean of Y for golfers who average 290 yards off the tee is $182.9-0.7394(290)+0.001203(290)^2 = 69.65$. Hence, $(Y|x = 290) \sim N(69.65, 0.25^2)$. The 10th percentile of the standard Normal distribution is -1.28. Therefore, the 10th percentile of the distribution of Y at $x = 290$ is:

$$69.65 - 1.28(0.25) = 69.33.$$

19. In the notes and in class, we discussed a simulation study that investigated how well AIC, BIC and adjusted-R^2 select the degree of a polynomial in regression. Which of the following is correct in regard to that study? (Read all options carefully.)

(a) AIC had a tendency to overestimate the true polynomial degree.

(b) BIC did a better job of selecting the true polynomial degree than the other two methods.

(c) Adjusted-R^2 had a strong tendency to overestimate the true polynomial degree.

(d) Both options (b) and (c) are correct.

(e) All of options (a), (b) and (c) are correct.

See pg. 67-68 of Chapter 1B notes. Since this was straight out of the notes and a fair warning was also given (and there was a similar question in the practice exam), you will **not** get any partial credit. You get credit only if you mark the *most* correct and inclusive option i.e. (e).

20. The method we discussed for identifying influential observations in regression is:

(a) The Sign test.

(b) The Kruskal-Wallis test.

(c) The Shapiro-Wilk test.

(d) The Cook's D statistic.

(e) The flop shot.

See pg. 110-111 of Chapter 1C notes.

21. The management of a museum wants to know if the tendency of visitors to buy souvenirs from the gift shop is dependent on the ages of the visitors. A total of 529 visitors were randomly selected from all of last year's visitors, and were cross-classified with respect to age and whether or not they bought a souvenir. The resulting data is presented below:

	Bought souvenir	Did not buy	Total
Under 18	61	108	169
18-50	93	137	230
Over 50	49	81	130
Total	203	326	529

Under the hypothesis that age and buying habits are independent of each other, what is an estimate of the expected number of visitors (out of 529) who are over 50 years old and do not buy from the gift shop?

(a) 49.

(b) 65.2.

(c) 80.1.

(d) 81.

(e) 141.7.

This corresponds to a contingency table based χ^2 test for independence of two characteristics (age and buying habits). Under H_0, using the formula from pg. 214-215 of Chapter 3 notes, the required expected number of counts for the $(3,2)^{th}$ cell is given by: $130(326)/529 = 80.1$.

22. In the setting of Problem **21**, the value of the χ^2 statistic for testing independence of age and buying habits is 0.810. Then, the correct conclusion for the test at a level of significance = 0.05, and an accurate bound for the P-value (P) of the test are:

(a) We conclude that buying habits depend on age, and $0.025 < P < 0.05$.

(b) We cannot conclude that buying habits depend on age, and $0.025 < P < 0.05$.

(c) We conclude that buying habits depend on age, and $0.10 < P < 0.90$.

(d) We cannot conclude that buying habits depend on age, and $0.10 < P < 0.90$.

(e) We have proven that buying habits do not depend on age, and $0.10 < P < 0.90$.

This corresponds to a contingency table based χ^2 test for independence of two characteristics (age and buying habits) with $I = 3$ and $J = 2$ categories (see Chapter 3). Under H_0, the test statistic χ^2 follows a $\chi^2_{(I-1)(J-1)} \equiv \chi^2_2$ distribution, and we reject H_0 for large values of χ^2. Hence, since its observed value is $\chi^2_{obs} = 0.81$, the P-value is $P(\chi^2_2 > \chi^2_{obs}) = P(\chi^2_2 > 0.81)$. Now, since $\chi^2_{2,0.1} = 0.211 < \chi^2_{obs} = 0.81 < \chi^2_{2,0.9} = 4.605$, the P-value must be between 0.1 and 0.9. This also shows that the P-value > 0.05. So, we fail to reject H_0 at $\alpha = 0.05$ and conclude that there is no significant evidence to indicate that buying habits depend on age.

23. A multiple linear regression analysis with four independent variables x_1, x_2, x_3 and x_4 is done in "R" and the `summary` command produced the following information.

Coefficients:

	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	-0.04026	0.04747	-0.848	0.397
x1	2.06662	1.42639	1.449	0.148
x2	-1.06572	0.79710	-1.337	0.182
x3	-3.01791	8.00507	-0.377	0.707
x4	-0.02623	0.04471	-0.587	0.558

Based on this information, it is fair to say that:

(a) Not all 4 of the independent variables are needed in the model.

(b) None of the independent variables is needed in the model.

(c) At least one of the four independent variables is needed in the model.

(d) All four of the independent variables are needed in the model.

(e) The COVID-19 pandemic has now truly engulfed the whole of the United States.

As we discussed several times in class, the strategy to use in this case is to delete the variable whose t-statistic has the largest P-value, assuming that the largest P-value is not smaller than, say, 0.10. So we would delete x_3 in this case, refit the model excluding this variable, and repeat the previous strategy. Because of possible correlation(s) between the independent variables, it is **not** a good strategy to delete all independent variables with large P-values. It **is** possible in this case that x_1, x_2 and x_4 are needed in the model. We also discussed these concepts and subtleties in detail while talking about the multicollinearity issue in class.

24. In a simple linear regression analysis, the test for $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ turned out to have a P-value of 0.047. Which of the following is the best interpretation of this P-value?

(a) There is a probability of only 0.047 that H_0 is true.

(b) There is a probability of 0.953 that H_a is true.

(c) If H_0 were true, then the probability that the test statistic would be equal to or more extreme than its observed value is 0.047.

(d) If H_a were true, then the probability that the test statistic would be equal to or more extreme than its observed value is 0.047.

(e) If H_0 were true, then the probability that the test statistic would be equal to or more extreme than its observed value is 0.953.

This was the right interpretation of P-values that I have discussed many times in class for various problems. Note: I also said that the P-value is **not** the probability that H_0 is true.

25. A psychologist investigates the difference in maze test scores for a strain of laboratory mice trained under different conditions. The experiment is conducted using 22 randomly selected mice of this strain. Six of the 22 were randomly assigned to a 'Control' group that received no training, eight were randomly assigned to be trained under 'Condition 1', and the rest were trained under 'Condition 2'. After training, each mouse was given a test score between 0 and 100 that reflected its performance in a test maze.

The type of experimental design used here was:

(a) A randomized block design.

(b) A crossover design.

(c) A completely randomized design.

(d) A split-plot design.

(e) One inspired by the Bauhaus school.

This is clearly an example of a completely randomized design (CRD); see pg. 161 of Chapter 2B notes. As discussed in class, a 1-way ANOVA is the right statistical analysis for a CRD.

26. The following (partly filled) ANOVA table was computed from the data in Question 25.

Source	df	Sum of squares	Mean square	F
Conditions		691.531		
Within conditions		1182.333		
Total	21	1873.864		

The psychologist wishes to test the null hypothesis that the average performance is the same for all three conditions. If she uses a level of significance = 0.05, then which of the following gives the correct values of the F-statistic and the F-cutoff value with which to compare her test statistic? (Here are some F-cutoff values that you may need to answer this question: $F_{3,19,0.05} = 3.13$; $F_{2,19,0.05} = 3.52$; $F_{2,20,0.05} = 3.49$; $F_{3,20,0.05} = 3.10$.)

(a) F-statistic = 5.556; F-cutoff value = 3.13.

(b) F-statistic = 3.704; F-cutoff value = 3.13.

(c) F-statistic = 5.556; F-cutoff value = 3.52.

(d) F-statistic = 3.704; F-cutoff value = 3.52.

(e) F-statistic = 3.704; F-cutoff value = 3.10.

The problem falls under a one-way ANOVA setup with $k = 3$ treatments (conditions) and a total sample size of $N = 22$. From the (partial) ANOVA table above, the sum of squares due to the treatments (SSTr) = 691.531 and sum of squares due to error (SSE) = 1182.333. Hence, using the formula from pg. 120-125 of Chapter 2A notes, the required F-statistic for testing the equality of means of the responses under the three conditions is given by:

$$F = \frac{MSTr}{MSE} = \frac{SSTr/(k-1)}{SSE/(N-k)} = \frac{(691.531/2)}{(1182.333/19)} = 5.556.$$

Under the null hypothesis, F follows a $F_{k-1,N-k} \equiv F_{2,19}$ distribution. Hence, for a test with level $\alpha = 0.05$, the correct percentile, i.e. the F-cutoff value, with which to compare the test statistic is: $F_{2,19,0.05} = 3.52$ (using the list of cut-offs given). Thus, (c) is the correct option.

27. I graded a quiz for 12 students in my class. Their scores (out of 10 points) are as follows:

$$7,\ 5,\ 5,\ 7,\ 7,\ 9,\ 8,\ 6,\ 9,\ 8,\ 6,\ 7.$$

But I realized I made a grading error and so I decided to add 1 point to each student's score.

Using this revised data, suppose I now use the Sign test to test the following hypotheses for the median (M) of the population of my students' scores: $H_0 : M = 7.5$ vs. $H_a : M \neq 7.5$. Then, the Z-statistic and the approximate P-value for this Sign test are closest to:

(a) Z-statistic = -1.1547; P-value = 0.125.

(b) Z-statistic = -1.1547; P-value = 0.75.

(c) Z-statistic = 1.1547; P-value = 0.75.

(d) Z-statistic $= 1.1547$; P-value $= 0.125$.

(e) Z-statistic $= 1.1547$; P-value $= 0.25$.

The revised data for the scores of the 12 students after adding 1 point to each score is:

$$8,\ 6,\ 6,\ 8,\ 8,\ 10,\ 9,\ 7,\ 10,\ 9,\ 7,\ 8.$$

The number of observations in this data that are greater than the null value 7.5 of the median is $Y = 8$. Using pg. 222-224 of Chapter 4 notes, the Z-statistic for the Sign test then equals:

$$Z = \frac{Y - n/2}{\sqrt{n}/2} = \frac{8 - 12/2}{\sqrt{12}/2} = \frac{2}{\sqrt{12}/2} = 1.1547.$$

Since the alternative H_a is two-sided, and $Z \sim N(0,1)$ under H_0, the P-value for this test is $2P(Z > |Z_{obs}|) = 2P(Z > 1.1547)$, where $Z \sim N(0,1)$ and $Z_{obs} = 1.1547$ is the observed value of the test statistic Z. Using the Normal table, this roughly equals $2(1 - 0.8749) = 0.25$.

28. A laboratory wishes to compare the effectiveness of two sunscreen products. 100 people volunteered to participate in an experiment designed by the lab. Each person had sunscreen A applied to one arm and sunscreen B applied to their other arm. The 100 subjects were then exposed to sunlight for a prescribed period of time before being examined for the degree of tanning on each arm. The lab could have applied sunscreen A to 50 persons and sunscreen B to the other 50. They chose to apply both sunscreens to each person instead because:

(a) Doing so makes the degrees of freedom for a t-test statistic larger.

(b) Doing so ensures that the differences between the people will have less effect on detecting a difference between the sunscreens.

(c) This makes the experimental results for sunscreen A and B independent of each other.

(d) This so makes the logistics simpler since we don't need to keep track of who gets which type of sunscreen.

(e) This ensures that all subjects will have grounds for a lawsuit when they get sunburned.

This relates to the comparison of paired sample vs. 2-sample settings which, as we discussed in class, correspond to special cases of an RBD and a CRD, respectively. In Chapter 2B, we discussed in detail the benefits of RBDs in reducing within group variability compared to CRDs. The same principle therefore applies to this situation as well. Hence (c) is correct.

```
> fit=lm(y~x1+x2+x3+x4)
> fit12=lm(y~x1+x2)
> fit23=lm(y~x2+x3)
> fit14=lm(y~x1+x4)

> summary(fit)

Call:
lm(formula = y ~ x1 + x2 + x3 + x4)

Residuals:
    Min      1Q  Median      3Q     Max 
-49.396 -13.384  -1.188  11.823  54.457 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  8.71778   19.08414   0.457    0.648    
x1           5.23047    0.30209  17.314   <2e-16 ***
x2          -3.44166    0.37161  -9.262   <2e-16 ***
x3          -0.01945    0.25754  -0.076    0.940    
x4           2.31569    1.43913   1.609    0.109    
---
Residual standard error: 20.73 on 195 degrees of freedom
Multiple R-squared:  0.8567,    Adjusted R-squared:  0.8538 
F-statistic: 291.4 on 4 and 195 DF,  p-value: < 2.2e-16

> anova(fit)
Analysis of Variance Table

Response: y
           Df Sum Sq Mean Sq   F value Pr(>F)    
x1          1 435690  435690 1013.9356 <2e-16 ***
x2          1  64091   64091  149.1531 <2e-16 ***
x3          1      4       4    0.0085 0.9267    
x4          1   1113    1113    2.5892 0.1092    
Residuals 195  83792     430                     
---

> summary(fit12)

Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max 
-53.090 -11.848  -1.162  12.081  56.250 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  11.9296    18.9933   0.628    0.531    
x1            5.1922     0.1571  33.056   <2e-16 ***
x2           -3.4910     0.2863 -12.194   <2e-16 ***
---
Residual standard error: 20.76 on 197 degrees of freedom
Multiple R-squared:  0.8548,    Adjusted R-squared:  0.8533 
F-statistic: 579.8 on 2 and 197 DF,  p-value: < 2.2e-16

> anova(fit12)
Analysis of Variance Table

Response: y
           Df Sum Sq Mean Sq F value    Pr(>F)    
x1          1 435690  435690  1010.9 < 2.2e-16 ***
x2          1  64091   64091   148.7 < 2.2e-16 ***
Residuals 197  84908     431                      
---
```

```
> summary(fit23)

Call:
lm(formula = y ~ x2 + x3)

Residuals:
    Min      1Q  Median      3Q     Max 
-83.683 -20.668  -1.358  22.768  81.655 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 128.3294    28.1482   4.559 9.01e-06 ***
x2           -6.6487     0.5115 -12.998  < 2e-16 ***
x3            3.7776     0.2122  17.799  < 2e-16 ***
---
Residual standard error: 32.89 on 197 degrees of freedom
Multiple R-squared:  0.6355,    Adjusted R-squared:  0.6318 
F-statistic: 171.7 on 2 and 197 DF,  p-value: < 2.2e-16

> anova(fit23)
Analysis of Variance Table

Response: y
           Df Sum Sq Mean Sq F value    Pr(>F)    
x2          1  28825   28825  26.645 5.952e-07 ***
x3          1 342745  342745 316.820 < 2.2e-16 ***
Residuals 197 213120    1082                      
---

> summary(fit14)

Call:
lm(formula = y ~ x1 + x4)

Residuals:
    Min      1Q  Median      3Q     Max 
-83.880 -17.343  -0.782  17.766  76.185 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -123.7013    20.5277  -6.026 8.12e-09 ***
x1             4.9875     0.2054  24.280  < 2e-16 ***
x4             3.5022     1.8883   1.855   0.0651 .  
---
Residual standard error: 27.26 on 197 degrees of freedom
Multiple R-squared:  0.7495,    Adjusted R-squared:  0.747 
F-statistic: 294.8 on 2 and 197 DF,  p-value: < 2.2e-16

> anova(fit14)
Analysis of Variance Table

Response: y
           Df Sum Sq Mean Sq  F value  Pr(>F)    
x1          1 435690  435690 586.1064 < 2e-16 ***
x4          1   2557    2557   3.4396 0.06515 .  
Residuals 197 146443     743                     
---
```