SPSS *t*-tests/ANOVA

Due at the Start of Lab: Paper 2

Rationale for Today's Lab Session

This tutorial is designed to ensure that you have a basic understanding of *t*-tests and Analysis of Variance (ANOVA). You will need these skills for Lab Assignment 6. These skills are also essential for academic and employment pursuits in research. Today, you will go through this tutorial with your lab instructor. You can work collaboratively on this tutorial but must work independently on the graded lab assignment.

Instructions

Warning
SPSS periodically changes the visual display and organization of menus. The instructions presented in this tutorial may need to be augmented marginally depending on the version of SPSS you are using. If you get stuck, use Google, or ask the lab instructor for help.

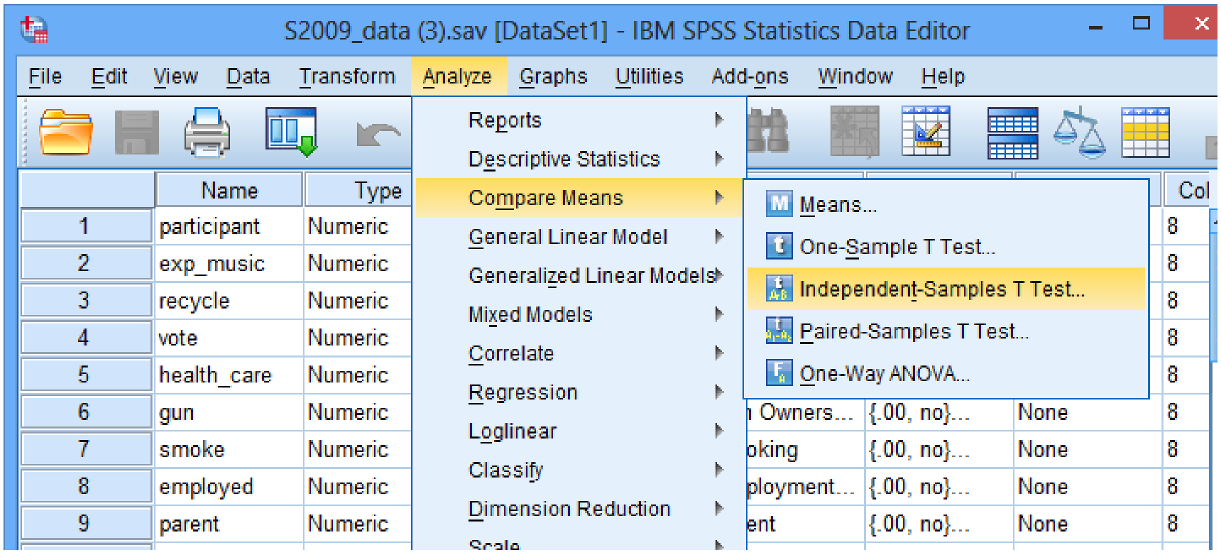
Accessing SPSS and Data
For the graded assignment, you will use the 2014 data files. However, for this tutorial, you will use the 2009 Pre-Existing Data files. Log on to BlackBoard, and download the 2009 files (S2009_data.sav, S2009_dictionary.xls). Open them in SPSS and Excel, respectively.

t-tests

For this section, you will learn to run a between-group t-test, the most commonly used type of *t*-test, and the second most common statistical analysis used in the social sciences (after the correlation).

Background Information. The between-group *t*-test is used when you have two groups or categories of people. It lets you see how the groups differ in terms of their scores on some continuous variable (a variable with an ordered rating system, like our 0-10 scales, age, etc.). The *t*-test is basically a means-to-an-end. It provides us with a *p*-value, the probability that a result is due to "chance" or "sampling error". The *p*-value shows up in the Output and tells us whether the result is likely due to chance. If $p < .05$, the difference between groups is reliable. If not, there is no reliable difference, and we tend to ignore the result.

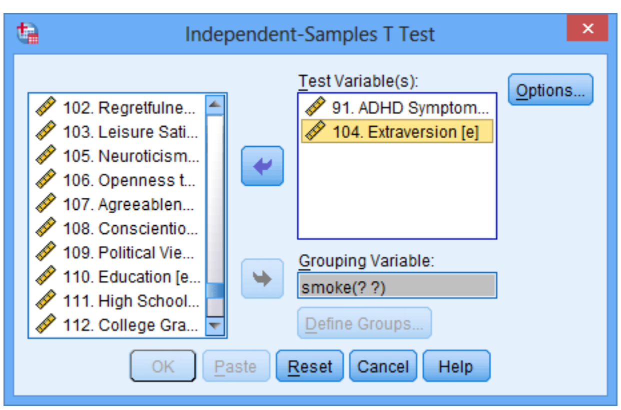
Running a *t*-test. Go to the Analyze menu, point to Compare Means, and choose "Independent-Samples T Test"

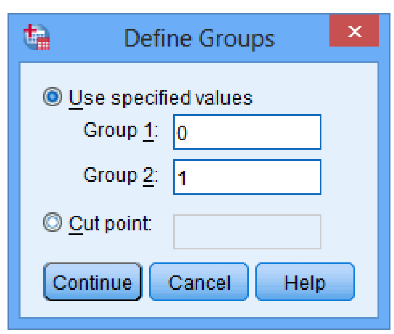


In the window that pops up, we always put the independent variable (grouping or categorical variable) in the "Grouping Variable" section of the box. In the "Test Variable(s)" box, put any continuous dependent variables you want to examine (you can choose more than one if you like). The analysis will tell us if the groups differ in terms of their scores on the "Test Variables".

Try putting Smoking (#6) in the "Grouping Variable" area, and put ADHD Symptoms (#91) and Extraversion (#104) in the "Test Variables" section, so we can see if smokers differ on these variables. At this point you will notice that the OK button is still gray, so we need to do one more step.

Single-click where it says "smoke(? ?)" in the Grouping Variables area. Then, click on the Define Groups button. SPSS needs you to tell it which numbers were used to describe the groups. In the data file, we arbitrarily coded nonsmoker = 0 and smoker = 1, so type a 0 where is says "Group 1" and a 1 where it says "Group 2". If you ever need to examine how variables were coded, you could use the Variable View option in SPSS or simply look in our Excel Data Guide file.





Click the Continue button, and then the OK button to run the analyses. Your Output should look something like this:

Group Statistics

	6. Smoking	N	Mean	Std. Deviation	Std. Error Mean
91. ADHD Symptoms	no	862	4.8979	2.25521	.07681
	yes	113	5.1681	2.64880	.24918
104. Extraversion	no	862	5.5882	1.91601	.06526
	yes	113	6.0442	2.21355	.20823

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
91. ADHD Symptoms	Equal variances assumed	8.860	.003	-1.172	973	.241	-.27023	.23051	-.72257	.18211
	Equal variances not assumed			-1.036	134.140	.302	-.27023	.26075	-.78594	.24548
104. Extraversion	Equal variances assumed	3.661	.056	-2.335	973	.020	-.45608	.19535	-.83944	-.07272
	Equal variances not assumed			-2.090	134.912	.038	-.45608	.21822	-.88766	-.02451

Using the top box, we see that smokers reported a marginally higher level of ADHD symptoms (*M* = 5.12, *SD* = 2.65) than did non-smokers (*M* = 4.88, *SD* = 2.56). The second box in the Output tells us whether the difference we observed in our sample was reliable. Based on the size of the difference and the number of people in our study, would we expect this difference to hold up for the population in general, or is it just a chance finding? The box tells us the *t*-value (-1.172), the degrees of freedom (a reference number, 973). If we weren't using a computer, we would compare this observed *t*-value to a critical *t*-value or cut score to determine whether the result is significant ($p < .05$). However, SPSS gives us the exact *p*-value, or the exact probability of obtaining this result by chance (.241). Basically, we'd expect to find this type of weak group difference just by chance alone 24% of the time. Because the *p*-value is greater than .05, the result is not statistically significant. The observed difference in our sample is not a reliable or trustworthy effect. At the population level, we would not typically expect smokers and non-smokers to differ in terms of ADHD symptoms.

In contrast, smokers (*M* = 5.12, *SD* = 2.65) scored higher than non-smokers (*M* = 5.12, *SD* = 2.65) on extraversion. The *t*-value (-2.335), degrees of freedom (973) and *p*-value (.020) are noted. For this analysis, $p < .05$, so the observed difference is trustworthy. At the population level, we would expect smokers to be more extraverted than non-smokers.

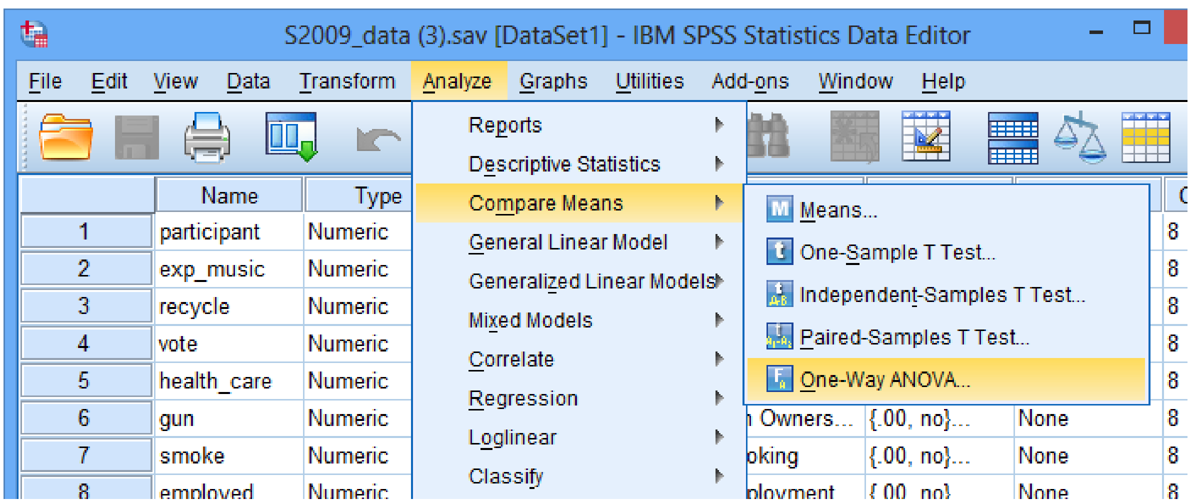
Practice Questions

1) Conduct a *t*-test examining whether choosing to Recycle (#2) is related to differences in ACT scores (#113). How would you report the result in APA style, using the APA Style Guide at the end of this assignment? Cohen's *d* can be calculated by hand using the formula from the notes (exp2.doc).
2) Conduct a *t*-test examining whether Support for Universal Health Care (#4) is related to differences in Racial Acceptance (#86). How would you report the result in APA style, using the APA Style Guide at the end of this assignment?
3) Conduct a *t*-test examining Gender (#11) differences in Political Views (#109). How would you report the result in APA style, using the APA Style Guide at the end of this assignment?

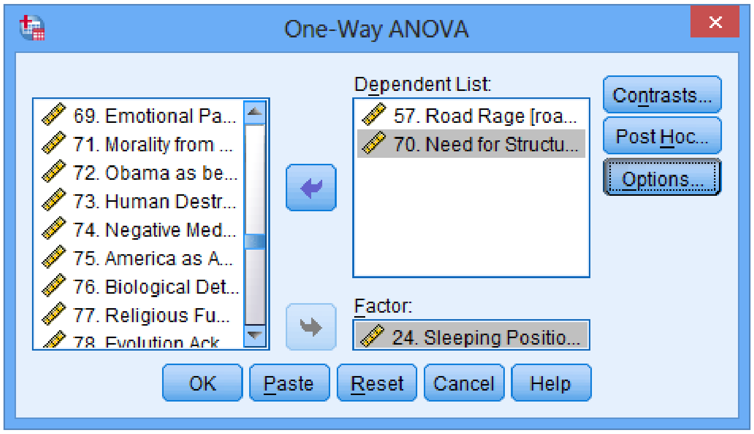
ANOVA

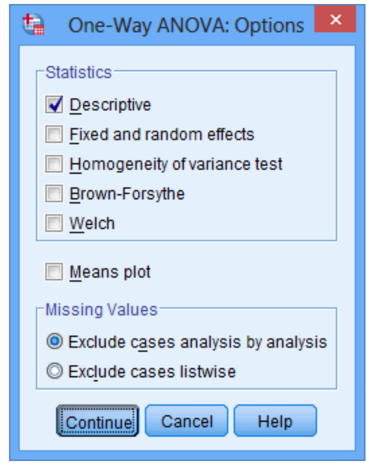
ANOVA is similar to the between-group *t*-test. It is used when there are more than two categories or groups of people. Again, these groups are compared in terms of their scores on some continuous variable.

Go to the Analyze menu, point to Compare Means, and choose One-Way ANOVA.



In the window that pops up, you put the categorical variable in the Factor area and the continuous dependent variable in the Dependent List area (feel free to include more than one). For practice, put Sleeping Position (#24) in the Factor area, and put Road Rage (#57) and Need for Structure (#70) in the Dependent List area. Before clicking OK, click on the Options button. Choose Descriptive in the pop-up window, telling SPSS to include basic descriptive statistics with our Output. Click Continue and then OK.





The Output should look something like this:

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
57. Road Rage	Back	110	4.2545	2.23991	.21357	3.8313	4.6778	1.00	9.00
	Stomach	267	4.7528	2.26421	.13857	4.4800	5.0256	1.00	9.00
	Side	491	4.2505	2.13627	.09641	4.0611	4.4399	1.00	9.00
	Fetal	107	5.0748	2.12665	.20559	4.6672	5.4824	1.00	9.00
	Total	975	4.4790	2.20027	.07047	4.3407	4.6173	1.00	9.00
70. Need for Structure	Back	110	7.1909	1.58835	.15144	6.8908	7.4911	1.00	9.00
	Stomach	267	7.2285	1.51069	.09245	7.0464	7.4105	1.00	9.00
	Side	491	7.3177	1.59107	.07180	7.1766	7.4588	1.00	9.00
	Fetal	107	7.5047	1.41003	.13631	7.2344	7.7749	1.00	9.00
	Total	975	7.2995	1.55017	.04965	7.2021	7.3969	1.00	9.00

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
57. Road Rage	Between Groups	89.172	3	29.724	6.239	.000
	Within Groups	4626.147	971	4.764		
	Total	4715.319	974			
70. Need for Structure	Between Groups	7.312	3	2.437	1.014	.386
	Within Groups	2333.238	971	2.403		
	Total	2340.550	974			

The top box is very detailed, so examine it slowly. The box tells how many people were in each group, and provides their basic descriptive statistics. For example, 107 people report sleeping in a fetal position, and as a group, they had the highest road rage scores (*M* = 5.07, *SD* = 2.13).

The second box in the SPSS Output provides the results of the ANOVA analysis, which are used to indicate whether there were any reliable differences across the groups of people. The *F*-value and degrees of freedom values are provided. If you are unfamiliar with these statistics from prior statistics courses, just be aware that they are used in the process of calculating a *p*-value. The *p*-value is listed in the table as ".000" due to rounding. [If you click on this value several times, a more exact value appears (e.g. 0.000339). The *p*-value can never be zero because there is always some probability that a result could be due to chance – in this case, the odds are very small.] Because $p < .05$, there are some reliable group differences present.

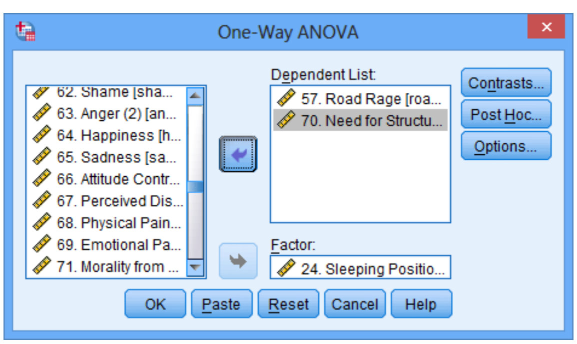
There were no significant group differences in Need for Structure because $p = .39$, which is obviously greater than .05. The minor mean differences in need for structure across groups are merely due to chance.

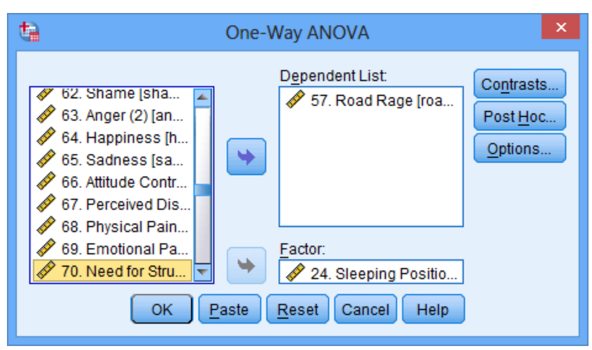
ANOVA Post-hoc Tests

The preceding ANOVA was simple to conduct, but it only provides limited information. The significantly *p*-value in the Road Rage example merely indicates that at least two of the groups reliably differed from each other. It could be that all four of the groups reliably differed from each other. It could be that only two differed from each other. Alternatively, one group might reliably differ from two other groups but not reliably differ from another group.

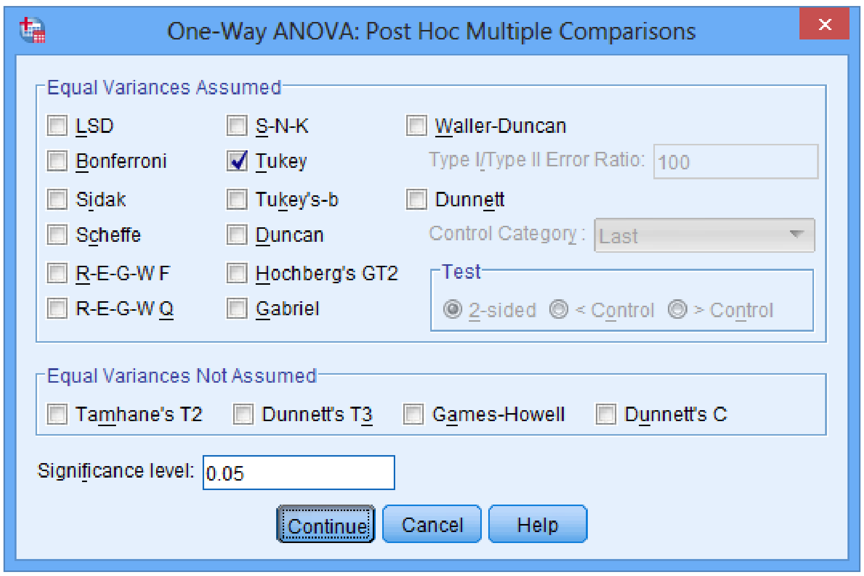
Look back at the red box in the Sleeping Position & Road Rage ANOVA Output. Do you think people who sleep on their Backs reliably differ in Road Rage from people who sleep on their Sides? Probably not. The difference is miniscule. Now compare Fetal vs. Back. These groups have the largest observed difference on Road Rage. Is that difference reliably? Probably. How about Stomach vs. Back? It's hard to say. The statistically significant ANOVA simply says that at least two groups differ – it is a signal that we need to conduct a follow-up test, called a "post-hoc test" to determine more specifically which groups reliably differ.

Post-hoc tests. We only conduct post-hoc tests if the ANOVA is statistically significant. Bring up the ANOVA window again (Analyze → Compare Means → One-Way ANOVA). The ANOVA involving Need for Structure (#70) was non-significant, so select it and click on the arrow in the middle of the box to remove it from the analyses.





Now, click on the “Post Hoc” button. Another window pops up. There are many different types of post-hoc tests. Some post-hoc tests are very liberal – they more easily declare a group difference to be statistically significant (e.g., Least Significant Difference, or “LSD”). Other post-hoc tests are more conservative – they are more cautious and tend to declare fewer group differences to be statistically significant (e.g., “Scheffe”). Many are somewhere in between. The LSD and Sheffe are very commonly used, and the most common in-between option is the “Tukey,” which is neither too liberal not conservative. Statisticians have and will continue to debate the merits of each post-hoc test. Acknowledging that that’s probably a very exciting debate, we’ll just stick with the Tukey for these analyses. Click Tukey, then Continue, then OK.



The Output should look familiar. It added two additional boxes.

Descriptives

57. Road Rage

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Back	110	4.2545	2.23991	.21357	3.8313	4.6778	1.00	9.00
Stomach	267	4.7528	2.26421	.13857	4.4800	5.0256	1.00	9.00
Side	491	4.2505	2.13627	.09641	4.0611	4.4399	1.00	9.00
Fetal	107	5.0748	2.12665	.20559	4.6672	5.4824	1.00	9.00
Total	975	4.4790	2.20027	.07047	4.3407	4.6173	1.00	9.00

ANOVA

57. Road Rage

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	89.172	3	29.724	6.239	.000
Within Groups	4626.147	971	4.764		
Total	4715.319	974			

Post Hoc Tests

Multiple Comparisons

57. Road Rage
Tukey HSD

(I) 24. Sleeping Position	(J) 24. Sleeping Position	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Back	Stomach	-.49826	.24730	.183	-1.1347	.1382
	Side	.00404	.23025	1.000	-.5885	.5966
	Fetal	-.82022*	.29638	.029	-1.5829	-.0575
Stomach	Back	.49826	.24730	.183	-.1382	1.1347
	Side	.50230*	.16597	.014	.0752	.9294
	Fetal	-.32196	.24974	.570	-.9647	.3207
Side	Back	-.00404	.23025	1.000	-.5966	.5885
	Stomach	-.50230*	.16597	.014	-.9294	-.0752
	Fetal	-.82426*	.23287	.002	-1.4236	-.2250
Fetal	Back	.82022*	.29638	.029	.0575	1.5829
	Stomach	.32196	.24974	.570	-.3207	.9647
	Side	.82426*	.23287	.002	.2250	1.4236

*. The mean difference is significant at the 0.05 level.

Homogeneous Subsets

57. Road Rage

Tukey HSD[a,b]

24. Sleeping Position	N	Subset for alpha = 0.05: 1	2
Side	491	4.2505	
Back	110	4.2545	
Stomach	267	4.7528	4.7528
Fetal	107		5.0748
Sig.		.157	.537

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 165.162.
b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

The third box of the Output provides the most important information. The Blue box compares the means and the Red oval provides *p*-values indicating which group differences are statistically significant. These tests are analogous to running t-tests on every possible two-group comparison.

The results are complicated, so it's good to have a general strategy for how to proceed. Usually, it's useful to begin with the highest-scoring group (in this case Fetal sleeper) and examine how they compare to each group. Then, look at how the second-highest group (Stomach sleepers) compares to each group with a lower mean, and so on.

- Fetal sleepers reliably have more Road Rage than people who sleep on their Backs (p = .03) or Sides (p = .002) but do not significantly differ from people who sleep on their Stomachs (p = .57).
- Stomach sleepers had more Road Rage than Side sleepers (p = .01) but did not differ from Back sleepers (p = .18), likely due to the lower sample size in that group.
- Back and Side sleepers do not differ (p > .99) [the table value says 1.00, but nothing can every reach 1.00. If you click on the value repeatedly, you'll see it's 0.999998.]

When there are many groups present (4, 5, 6…), it's easy for beginners to get overwhelmed by these post-hoc tests, which is why it's useful to use the above procedure of working from highest-to-lowest. It's also useful to end by summarizing the results as simply as possible. "The ANOVA indicated that sleeping position was significantly related to road range. Post-hoc analyses revealed that fetal sleepers had higher road rage than most other groups, back and side sleepers were lower than most other groups, and stomach sleepers were somewhere in between."

Practice Questions

4) Conduct an ANOVA analysis examining whether favorite Movie Genre (#23) is related to differences on the second Life Satisfaction variable (#60). How would you report the result in APA style, using the APA Style Guide at the end of this assignment? Cohen's *d* calculations are optional, though sometimes helpful for writing up ANOVA results.
5) Conduct an ANOVA analysis examining whether one's Favorite Season (#22) differences in Cell Phone Use (#55). How would you report the result in APA style, using the APA Style Guide at the end of this assignment?
6) Conduct an ANOVA analysis examining whether one's favorite Food Choice (#17) differences in Emotional Pain (#69). How would you report the result in APA style, using the APA Style Guide at the end of this assignment? Is it appropriate to examine the Post-Hoc results at all?

Choosing Tests

A number of statistical tests have been covered in this course: correlation, regression, between-group *t*-test, and ANOVA. The guide below provides input on when to use each test.

Independent Variable(s)	Dependent Variable	Test
1 dichotomous variable (1 variable with 2 categories)	1 continuous variable	t-test [a b]
1 polytomous variable (1 variable with more than 2 categories)	1 continuous variable	ANOVA [b]
1 continuous variable	1 continuous variable	Correlation [b]
Several dichotomous or continuous variables	1 continuous variable	Multiple Regression [b]

[a] Could use a correlation under those circumstances, though the *t*-test is generally preferred

[b] Could conduct these analyses when the dependent variable is dichotomous. Technically, other tests, such as chi-square and logistic regression are more appropriate, but no need to worry about those for this class.

Practice Questions

7) What statistical analysis would be most appropriate for examining the relationship between the following variables? Sadness (#66) to predict Crying (#52)
8) What statistical analysis would be most appropriate for examining the relationship between the following variables? Gender (#11) to predict Tanning (#45)
9) What statistical analysis would be most appropriate for examining the relationship between the following variables? Extraversion (#104) and Neuroticism (#105) to predict Leadership (#90)
10) What statistical analysis would be most appropriate for examining the relationship between the following variables? perceived Best Trait (#19) to predict Somatization (#40)

<u>APA Style Guide</u>
Note: You have my permission to copy any or all of this writing for this or future assignments.

Correlation Only (Significant, $p < .05$):

Example 1: The correlation between IQ and hours of television watched was significant, $r = -.35$, $p = .02$. That is, people who were smarter watched moderately less television.

Example 2: The correlation between IQ and hours of television watched was significant, $r = -.35$, $p < .05$. That is, people who were smarter watched moderately less television.

Include the correlation. When significant, say "$p < .05$" or provide the exact *p*-value. Then describe the results in plain English.

Correlation Only (Non-Significant, $p > .05$):

Example 1: IQ and number of hours of television watched were not significantly related, $r = .08$, $p = .67$. Thus, one's level of intelligence was not related to time spent watching television.

Example 2: IQ and number of hours of television watched were not sizably related, $r = .08$, *ns*. Thus, one's level of intelligence was not related to time spent watching TV.

Include the correlation. When non-significant, say "ns" for non-significant, or include the exact p-value. Then describe the results in plain English.

Several Correlations, followed by Multiple Regression:

Example 1: Family stress ($r = .48$, $p < .05$), work stress ($r = .56$, $p < .05$), and school stress ($r = .21$, $p < .05$) all significantly predicted overall life stress. However, social support did not predict level of life stress, $r = .03$, *ns*. Thus, although social support was not related to life stress, one's level of school stress was slightly related, family stress was modestly related, and work stress was strongly related to level of life stress. To examine the overall contribution of the three significant predictors (school stress, family stress, and life stress) in accounting for life stress, multiple regression was used. The results of the multiple regression analysis indicate that these three predictors accounted for a large proportion of the variance in life stress, $R^2 = .40$, $p < .05$. Thus, school stress, family stress, and work stress together account for 40% of the differences in overall life stress.

Example 2: Several factors were hypothesized to predict college GPA. Being encouraged to read ($r = .19$, $p = .002$) and conscientiousness ($r = .26$, $p < .001$) had small positive relationships with college GPA. ADHD symptoms had a small negative relationship ($r = -.17$, $p = .007$). Hours of work per week was not correlated with GPA ($r = .08$, $p = .22$). Thus, being encouraged to read and being conscientious are related to better grades, but having ADHD symptoms is related to lower grades. The number of hours people spend on employment was not related to grades. Multiple regression was used to examine the combined effect of being encourages to read, conscientiousness, and ADHD symptoms on college GPA. These three predictors combined to modestly predict GPA, R = .33, R^2 = .11, p < .001. Therefore, being encouraged to read, conscientiousness, and ADHD symptoms explain 11% of the differences in college grades.

First, describe the correlational results, where you compare each of the predictors to the dependent variable. Then, provide a rationale for the regression analyses. In reporting the results, people usually include R, R^2, or both, followed by the p-value. Then, describe the results in plain English.

t-test (Significant, *p* < .05):

Example: Females (M = 2.34, SD = 2.06) tan slightly more often than males (M = 1.60, SD = 1.46), which was a significant effect, d = .42, t(298) = 3.11, p = .002. Thus, women are more likely to go tanning than men.

Be sure to include the basic descriptives (M and SD), d (calculated by hand), t, df (in parentheses), and the p-value. You may list $p<.05$ instead of the exact p-value if you like.

t-test (Non-Significant, *p* > .05):

Example: Smokers (M = 5.01, SD = 2.38) were slightly moodier than non-smokers (M = 4.50, SD = 2.18); however, this differences was non-significant, d = .26, t(298) = 1.64, p = .10. That is, smoking is unrelated to moodiness.

Be sure to include the basic descriptives (M and SD), Cohen's d (calculated by hand), t, df (in parentheses), and the p-value. You may list *ns* instead of the exact p-value if you like.

ANOVA (Significant, *p* < .05):

Example: Music device preference was significantly related to openness to experience, F(2,297) = 4.30, p = .02. People who listen to vinyl or cassettes were highest (M = 8.33, SD = 0.78) on openness to experience, followed by .mp3 listeners (M = 6.91, SD = 1.75), followed by CD listeners (M = 6.86, SD = 1.66). A post-hoc test revealed that each of the group differences was statistically significant. People who use older music devices are more open.

Include the degrees of freedom (the top two df values in the Output), the F-value, and the p-value. Include basic descriptive statistics as well. You may calculate out a Cohen's d to compare any two of the specific groups, if you so choose.

ANOVA (Non-Significant, *p* > .05):

Example: Drivers, walkers, and bikers did not differ significantly in terms of religious involvement, F(2,297) = 1.65, p = .19. Thus, transportation mode is not related to involvement with religion.

Keep the results concise. Include the degrees of freedom (the top two df values in the Output), the F-value, and the p-value. You may list *ns* instead of the exact p-value if you like.