Statistics - Math 222 **Midterm 2 Pretest**

Name: _____ **Instructor: Lins**

This is not for a grade! This is just to test yourself and see what you know and what you need to study for next week.

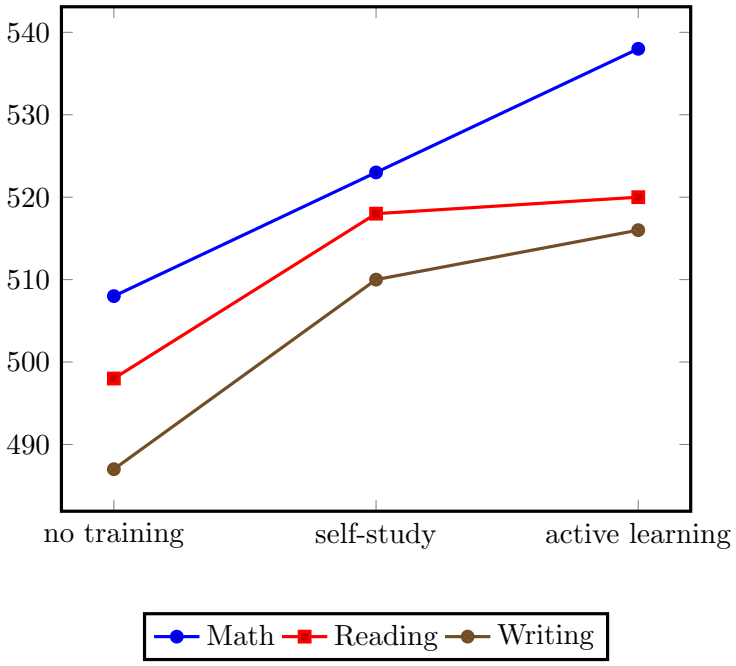
1. In each case, state the specific statistical procedure that is appropriate for the given situation. Be specific: identify the response variable and the explanatory variable(s). If there are any categorical variables present, state how many levels each categorical variable has.

   (a) Though most people are aware of the benefits of regular exercise, those who do not enjoy exercise are less likely to work out regularly. A study was designed to determine whether exercise enjoyment could be accurately predicted from certain survey data. The subjects were 282 female participants in an organized exercise class. Exercise enjoyment was measured on a scale from 18 to 136. The factors considered were (a) satisfaction with the music used in the class (on a scale from 4 to 28), (b) satisfaction with the instructor (on a scale from 6 to 42), and (c) self-concept, a variable that measured the extent to which the subject viewed herself as an exerciser (on a scale from 0 to 10). The results of the study showed a statistically significant linear association between the variables.

   (b) A study examined the effects of political orientation and religious outlook on the amount of charitable giving a person contributes in a year. Subjects were self-classified as either conservative, moderate, or liberal, and as either very religious, somewhat religious, neutral, somewhat atheistic, or very atheistic.

   (c) The developer of a virtual reality (VR) teaching tool for the deaf wants to compare the effectiveness of different navigation methods. A total of 40 children were available for the experiment, of which equal numbers were randomly assigned to use a joystick, wand, dance mat, or gesture-based pinch gloves. The time (in seconds) to complete a designed VR path is recorded for each child.

2. An experiment was conducted to compare the efficacy of two SAT preparation techniques. One group of subjects was given a course of self-study materials. A second group was given an on-line active-learning course in which participants could call on the instructor for help. As a control, a third group of subjects was given no training. After the two training courses were completed, all subjects were given the SAT test, and their scores were recorded separately for the reading, writing, and mathematics sections of the test. Below is a plot of the mean scores versus the training method, for each of the three sections of the SAT.



Use this plot to answer the following questions.

(a) Describe specifically the design of this experiment.

(b) The two factors in this experiment are "training type" and "SAT section". Describe specifically the main effect of each of the factors in this experiment.

(c) Describe specifically any interaction between the factors.

(d) The P-value for the main effect due to SAT section was P = 0.23. Explain specifically what this means.

3. A driver calculates the fuel economy of each tank of gas by dividing the mileage on the odometer by the amount of gas he puts in the car. The car also has a built in computer that estimates fuel economy (in MPG). When comparing the two different fuel economy calculations over 20 tanks of gas, the driver finds that in 17 out of the 20 tanks, the computer gave a higher estimate than the driver's own calculations. Carry out a simple sign test of the hypotheses:

   H_0: The median difference between the two calculations is zero.
   H_A: The median difference is not zero.

   You can use either the exact binomial distribution or the normal approximation to the binomial distribution with a continuity correction.

4. If we had access to all of the fuel economy data from the last problem, would it be appropriate to use a Wilcoxon signed rank test for matched pairs to test the hypotheses in the last problem? Explain why or why not.

5. Which of the following is not an advantage of two-way ANOVA over one-way ANOVA.

   A. Two-way ANOVA lets you test two factors at once.
   B. Two-way ANOVA is less prone to bias than one-way ANOVA.
   C. Two-way ANOVA lets you detect interactions between two factors.
   D. Two-way ANOVA is more powerful than one-way ANOVA.

6. Do piano lessons improve the spatio-temporal reasoning of children? The data below shows the change in spatio-temporal reasoning of 34 children who took piano lessons, 10 who took singing lessons, 20 who had some computer instruction, and 14 who received no extra lessons.

Lessons	n	$\bar{x}$	SE
Piano	34	3.6176	3.0552
Singing	10	-0.3000	1.4944
Computer	20	0.4500	2.2118
None	14	0.7857	3.1908

The ANOVA table for these results is:

	Df	SS	MS	F	p
Lesson	3	207.28	69.093	9.6129	1.953e-05
Residuals	74	553.43	7.1875		
Total	77	760.71	9.8794		

(a) What is are the null hypothesis and alternative hypothesis for this ANOVA test?

(b) What can you conclude based on the ANOVA table above?

(c) In particular, we wanted to know if there was a difference between kids who took piano lessons and the other kids in spatio-temporal reasoning improvement. Compute a 95% confidence interval for the contrast that compares the mean of the piano lesson group with the average of the other three means.

Recall that the standard error of a contrast is

$$SE_c = s_p \sqrt{\sum \frac{a_i^2}{n_i}}$$

where $s_p = \sqrt{MSE}$ *is the pooled sample standard deviation,* n_i *are the sample sizes for each group, and* a_i *are the coefficients of the contrast.*

7. In sedentary individuals, low blood pressure (hypotension) often occurs after a single bout of aerobic exercise and lasts nearly two hours. This can cause dizziness, light-headedness, and possibly fainting upon standing. It is thought that endurance training can reduce the degree of postexercise hypotension. To test this, researchers studied 16 endurance-trained and 16 sedentary men and women. The following table summarizes the postexercise systolic arterial pressure (mmHg) after 60 minutes of upright cycling.

Group	n	$\bar{x}$	SE
Women, sedentary	8	100.7	3.4
Women, endurance	8	105.3	3.6
Men, sedentary	8	114.2	3.8
Men, endurance	8	110.2	2.3

(a) Make an interaction plot for this data. Describe the pattern that you see.

(b) From the table, one can show that $SSA = 677.12$, $SSB = 0.72$, $SSAB = 147.92$, and $SSE = 2478$, where A is the gender effect, and B is the training level. Construct the ANOVA table with degrees of freedom, mean squares, and F-statistics, for each effect (You don't need to compute p-values).

(c) Does it look like the assumption of constant variance is statisfied? If so, what is the best estimate for σ^2?

t-Distribution Critical Values (Table C)

Confidence Level	60%	80%	90%	95%	98%	99%	99.8%	99.9%
dF								
1	1.376	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.920	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.906	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.889	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.879	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.876	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.873	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.870	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.868	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.866	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.865	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.863	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.862	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.861	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.860	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.859	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.858	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.858	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.857	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.856	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.856	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.855	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.855	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.854	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.854	1.310	1.697	2.042	2.457	2.750	3.385	3.646
31	0.853	1.309	1.696	2.040	2.453	2.744	3.375	3.633
32	0.853	1.309	1.694	2.037	2.449	2.738	3.365	3.622
33	0.853	1.308	1.692	2.035	2.445	2.733	3.356	3.611
34	0.852	1.307	1.691	2.032	2.441	2.728	3.348	3.601
35	0.852	1.306	1.690	2.030	2.438	2.724	3.340	3.591
36	0.852	1.306	1.688	2.028	2.434	2.719	3.333	3.582
37	0.851	1.305	1.687	2.026	2.431	2.715	3.326	3.574
38	0.851	1.304	1.686	2.024	2.429	2.712	3.319	3.566
39	0.851	1.304	1.685	2.023	2.426	2.708	3.313	3.558
40	0.851	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	0.849	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	0.848	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.846	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.845	1.290	1.660	1.984	2.364	2.626	3.174	3.390
500	0.842	1.283	1.648	1.965	2.334	2.586	3.107	3.310
z^*	0.842	1.282	1.645	1.960	2.326	2.576	3.090	3.291
One-sided p-value	0.2	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
Two-sided p-value	0.4	0.2	0.1	0.05	0.02	0.01	0.002	0.001