Extra Questions

Q1. Which statistical methods do we use for paired categorial data?

A. 2x2 contingency table / t-test

B. McNemar's test

C. Fisher's exact test

D. Z-test

Extra Questions

Q2. When do we use Fisher's Exact test for categorial data?

A. When the expected values of 2x2 table are small

B. When the observed values of 2x2 table are small

D. When the expected values of 2x2 table are large

C. When the observed values of 2X2 table are large

Extra Questions

Q3. The test statistic for testing the relation between two discrete variables (one with R categories and the other with C categories, i.e. RxC table) follows what kind of distribution:

A. t-distribution

B. normal distribution

C. Chi-square distribution

D. Binomial distribution

Extra Questions

Q4. What is the expected values for:

1) Blonde hair and blue eye

2) Black hair and brown eye

A. 1.64 and 3.64

B. 1.64 and 5

C. 5 and 3.64

D. 3.64 and 5

	Eye Color				
Hair Color	Green	Blue	Brown	Black	Total
Blonde	4	7	2	1	14
Brown	2	4	18	2	26
Black	1	2	5	2	10
Total	7	13	25	5	50

Extra Questions

Q5. Which statistical methods do we use for the data shown in the table?

TABLE 1. Detection of SARS CoV by quantitative PCR and real-time LAMP assays

Quantitative RT-PCR	RT-LAMP		Total
	Positive	Negative	
Positive	40	6[a]	46
Negative	2	11	13
Total	42	17	59

[a] Three of these samples were positive in a repeated LAMP test.

A. 2x2 contingency table / t-test

B. McNemar's Chi-square test (Normal-theory test)

C. McNemar's exact test

D. Z-test

Extra Question

Q1. A study investigates the association between the size of school and academic performance to examine if the size of the school is linked with academic performance. "api00" is the dependent variable and "enroll" is the independent variable (predictor). Let's say the estimated regression line: api00= a + b *enroll

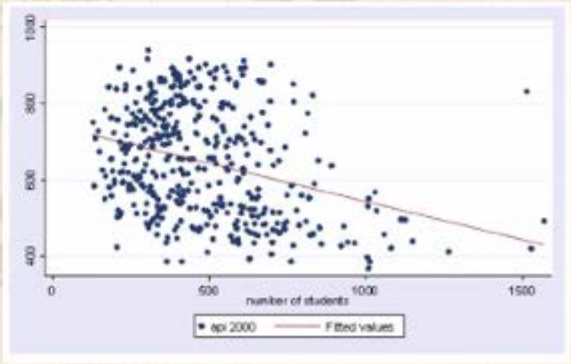
From the figure, the coefficient (b) should be:

A. 0

B. Positive

C. Negative

D. Infinity



Extra Question

Q2. Calculate the predicted average API score of a class size of 500 students.

A. 644.32
B. 644.318
C. 532.33
D. 532

Source	SS	df	MS	Number of obs =	400
Model	817326.293	1	817326.293	F(1, 398) =	44.83
Residual	7256345.70	398	18232.0244	Prob > F =	0.0000
				R-squared =	0.1012
Total	8073672.00	399	20234.7669	Adj R-squared =	0.0990
				Root MSE =	135.03

api00	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
enroll	-.1998674	.0298512	-6.70	0.000	-.2585532	-.1411817
_cons	744.2514	15.93308	46.71	0.000	712.9279	775.5749

Extra Question

Q3. Which of the following statement about the F test is incorrect?

Source	SS	df	MS	Number of obs =	395
Model	6740702.01	9	748966.89	F(9, 385) =	232.41
Residual	1240707.78	385	3222.61761	Prob > F =	0.0000
				R-squared =	0.8446
				Adj R-squared =	0.8409
Total	7981409.79	394	20257.3852	Root MSE =	56.768

api00	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ell	-.8600707	.2106317	-4.08	0.000	-1.274203	-.4459382
meals	-2.948216	.1703452	-17.31	0.000	-3.28314	-2.613293
yr_rnd	-19.88875	9.258442	-2.15	0.032	-38.09218	-1.68531
mobility	-1.301352	.4362053	-2.98	0.003	-2.158995	-.4437089
acs_k3	1.3187	2.252683	0.59	0.559	-3.1104	5.747801
acs_46	2.032456	.7983213	2.55	0.011	.462841	3.602071
full	.609715	.4758205	1.28	0.201	-.3258169	1.545247
emer	-.7066192	.6054086	-1.17	0.244	-1.89694	.4837018
enroll	-.012164	.0167921	-0.72	0.469	-.0451798	.0208517
_cons	778.8305	61.68663	12.63	0.000	657.5457	900.1154

A. The 9 variables, when considered together, are significant predictors of API score.

B. At least one of the coefficients of the multiple regression model is not equal to 0.

C. The degree of freedom = 385.

D. Coefficient of "ell" is not equal to zero when the coefficients of other variables are not equal to zero.

Extra Question

Q4. Which are the statistically significant predictors of API score after holding other variables constant?

A. ell, meals, yr_rnd, mobility, acs_k3, acs_46, full, emer, enroll

B. ell, meals, yr_rnd, mobility, acs_k3, acs_46, full, emer

C. ell, meals, yr_rnd, mobility, acs_46

D. Not sure

Source	SS	df	MS		
Model	6740702.01	9	748966.89	Number of obs	= 395
Residual	1240707.78	385	3222.61761	F(9, 385)	= 232.41
				Prob > F	= 0.0000
Total	7981409.79	394	20257.3852	R-squared	= 0.8446
				Adj R-squared	= 0.8409
				Root MSE	= 56.768

api00	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ell	-.8600707	.2106317	-4.08	0.000	-1.274203	-.4459382
meals	-2.948216	.1703452	-17.31	0.000	-3.28314	-2.613293
yr_rnd	-19.88875	9.258442	-2.15	0.032	-38.09218	-1.68531
mobility	-1.301352	.4362053	-2.98	0.003	-2.158995	-.4437089
acs_k3	1.3187	2.252683	0.59	0.559	-3.1104	5.747801
acs_46	2.032456	.7983213	2.55	0.011	.462841	3.602071
full	.609715	.4758205	1.28	0.201	-.3258169	1.545247
emer	-.7066192	.6054086	-1.17	0.244	-1.89694	.4837018
enroll	-.012164	.0167921	-0.72	0.469	-.0451798	.0208517
_cons	778.8305	61.68663	12.63	0.000	657.5457	900.1154

Extra Question

Q5. **When one or both of the variables are either ordinal or have a distribution that is far from normal, which of the following test should be used to assess the relation between two variables?**

A. Pearson's correlation
B. Spearman rank correlation
C. Single linear regression
D. Multiple linear regression

Extra Question

Q1. A researcher was interested in stress levels of students during lectures. He took the same group of 8 students and measured their anxiety during a normal lecture and again in a lecture in which he had peers to be disruptive and misbehave. The data were not normally distributed. Which test should he use to compare the experimental conditions?

A. Z-test
B. Paired samples t-test
C. Wilcoxon rank-sum test
D. Wilcoxon signed-rank test

Extra Question

Q2. A researcher measured 20 people's physiological reactions while watching comedy movies and compared them to 30 people's physiological reactions when watching horror movies. The data were skewed. What test should be used to analyse the data?

A. Independent t-test

B. Dependent (related) t-test

C. Wilcoxon rank-sum test

D. Wilcoxon signed-rank test

Extra Question

Q3. Which of the following is true about nonparametric statistical methods?

A. All of the following are true

B. when no assumptions about the shape of the distribution

C. when we cannot apply central-limit theorem due to small sample size

D. when the data are skewed / not normally distributed.

Extra Question

Q4. A new treatment is proposed for patients with lung cancer. Investigators are concerned with patient's ability to tolerate the treatment and assess their quality of life both before and after receiving the new treatment. Quality of life (QOL) is measured on an ordinal scale and for analysis purposes, numbers are assigned to each response category as follows: 1=Poor, 2= Fair, 3=Good, 4= Very Good, 5 = Excellent. The data are shown below.

If we would like to test whether there is a difference in QOL after the treatment as compared to before. What kind of test should we use?

Patient	QOL Before Chemotherapy Treatment	QOL After Chemotherapy Treatment	Difference (After-Before)	Sign
1	3	2	-1	-
2	2	3	1	+
3	3	4	1	+
4	2	4	2	+
5	1	1	0	-
6	3	4	1	+
7	2	4	2	+
8	3	3	0	+
9	2	1	-1	-
10	1	3	2	+
11	3	4	1	+
12	2	3	1	+

A. Wilcoxon rank-sum test

B. Dependent (related) t-test

C. Sign test

D. Wilcozon signed-rank test

Extra Question

Q5. Which of the following shows the correct matching:

Analysis Type	Parametric Procedure	Nonparametric Procedure
Compare means between two distinct/independent groups	Two-sample t-test	1
Compare two quantitative measurements taken from the same individual	Paired t-test	2
Estimate the degree of association between two quantitative variables	Pearson coefficient of correlation	3

A. 1: Sign test, 2: Wilcoxon signed-rank test, 3: Spearman's rank correlation

B. 1: Wilcoxon rank-sum test, 2: Wilcoxon signed-rank test, 3: Spearman's rank correlation

C. 1: Wilcoxon rank-sum test, 2: sign test, 3: Spearman's rank correlation

D. 1: Wilcoxon rank-sum test, 2: Wilcoxon signed-rank test, 3: Pearson's correlation