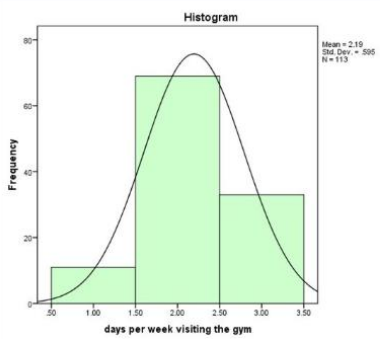
W1

Which of the two graphs is suitable to describe the variable created from the question

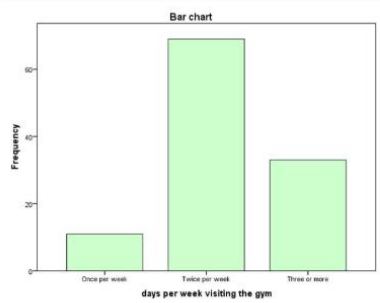
How many times do you go to the gym? a) once, b) twice, c) three or more

Select one:

a.



b.



Based on the below frequency table for 'gender' which of the following interpretations is correct?

gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	48	40.0	40.7	40.7
	female	64	53.3	54.2	94.9
	non-binary	6	5.0	5.1	100.0
	Total	118	98.3	100.0	
Missing	System	2	1.7		
Total		120	100.0		

Select one:

a. 54.2% of the people who responded identified themselves as female

b. 59.3% of people who responded identified themselves as female

c. 94.9% of the people in the sample identified themselves as female

d. 54.2% of the people in the sample identified themselves as female ×

Even though we cannot see the histogram for the 'age' variable, we may expect 'age' to be normally distributed because?

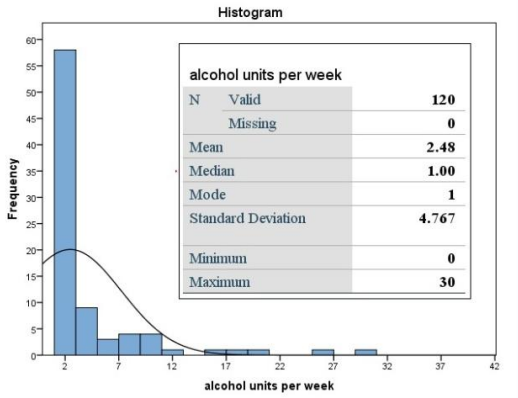
Statistics

age

N	Valid	119
	Missing	1
Mean		24.7370
Median		24.5000
Mode		24.00
Std. Deviation		1.95413
Minimum		19.70
Maximum		29.60
Percentiles	25	23.5000
	50	24.5000
	75	26.2000

Select one:

- a. The mean, median and mode are close together. ✓
- b. The minimum value is less than 10 units smaller than the maximum value.
- c. The standard deviation is small.
- d. The inter-quartile range is narrow



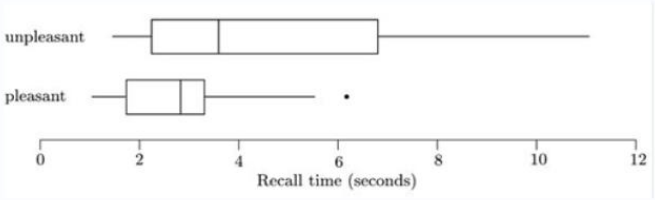
Select one:

- a. The people in the sample consumed on average 1 units of alcohol per week (SD= 4.77)
- b. Half the people in the sample consumed less than 2.48 units of alcohol per week.
- c. Half the people in the sample consumed less than 1 unit of alcohol per week (min=0, max = 30). ✓
- d. The people in the sample consumed on average 2.5 units of alcohol per week (min=0, max = 30).

Which of the following lists all parts of the summary for a box plot?

Select one:

- a. Smallest, Largest, Mean, Standard Deviation and Variance
- b. Mean, Median, Mode, Range, and Standard Deviation
- c. Minimum, Quartile 1, Median, Quartile 3, and Maximum ✓
- d. Minimum, Maximum, Range, Mean, and Median



Select one:

a. Both box plots are normally distributed.

b. The unpleasant category has a high value outlier

c. The dispersion of recall times is larger for the ‘unpleasant’ sample than the 'pleasant' sample.

d. Recall times for 'unpleasant' memories are shorter than those for 'pleasant' memories.

W2

Which of the below statements is the correct definition of the sampling distribution?

Select one:

- a. The distribution of the estimated values for a parameter, based on random samples of the same size from a population ✓
- b. The distribution of data in one sample
- c. The distribution of the parameter in a population
- d. The distribution of the values of the statistic for all individuals in one sample

Which of the statements below is correct?

Select one:

- a. The mean of the sampling distribution is equal to the mean of the population divided by the sample size.
- b. The mean of the sampling distribution is equal to the mean of the population divided by the square root of the sample size.
- c. The SD of the sampling distribution is equal to the SD of the population divided by the square root of the sample size ✓
- d. The SD of the sampling distribution is equal to the SD of the population divided by the sample size.

Which of the statements below is correct?

Select one:

- a. The 'statistic' sample mean provides the point estimate of the population 'parameter' mean. ✓
- b. The 'parameter' sample mean is the estimated value of the population 'statistic' mean.
- c. The 'parameter' sample mean is a point estimate of the population 'statistic' mean.
- d. The 'statistic' sample mean is always equal to the population 'parameter' mean.

A researcher estimated a 95% CI to be [17.8, 22.2]. If the researcher calculated the 99% CI, this would be:

Select one:

- a. Wider [17.1, 22.9]. ✓
- b. Narrower [18.2, 21.8].

Two researchers used two different samples and studied the expected hours of sleep in the population. Both samples had the same mean 8.5 and standard deviation of 2 hours. The 95% confidence interval for Researcher A was [7, 9] and the 95% confidence interval for Researcher B was [6, 11]. Which of the two researchers had the larger sample size?

Select one:

- a. Researcher B
- b. Researcher A ✓

W3

A researcher wishes to test if the mean hours of work in a population is 40h per week. She samples N=100 individuals and finds a sample mean of 45h per week. Which of the below is the correct null hypothesis.

Select one:

a. H_0: The population mean is different than μ_0= 45 hours/week

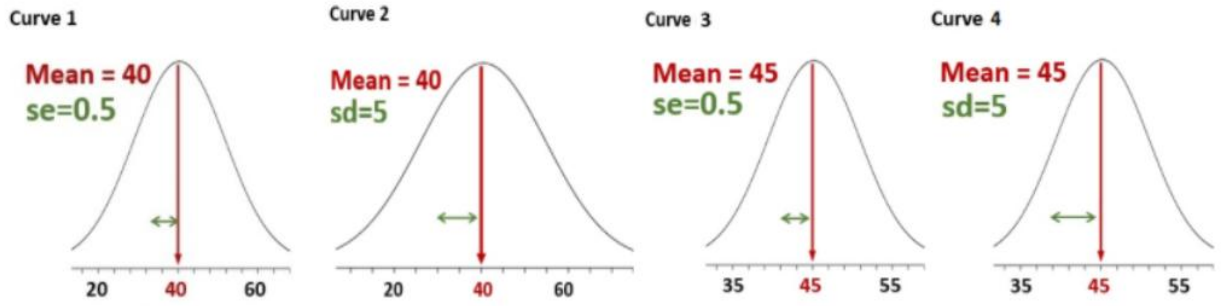
b. H_0: The population mean is equal to μ_0= 40 hours/week ✓

c. H_0: The population mean is equal to μ_0= 45 hours/week

d. H_0: The population mean is different than μ_0= 40 hours/week

A researcher tests if the mean hours of work in a population is μ_0=40 hours per week. He samples N=100 individuals and finds a sample mean of 45h/week and a standard deviation of 5 hours per week. The estimated sample error is therefore 0.5 hours per week.

Which of the following normal curves represents the sampling distribution under the null hypothesis?



Select one:

1. Curve 1 ✓

A researcher uses a one sample t-test to test the null hypothesis: 'the population mean equals 40h per week'.

The result shows that the p-value = 0.089. What does the researcher need to state in his report?

Select one:

a. We reject the null hypothesis that the population mean is 40h/week

b. We do not reject the null hypothesis that the population mean is 40h/week ✓

c. We accept the null hypothesis that the population mean is 40h/week

d. We accept the alternative hypothesis that the population mean is different to 40h/week

The p-value is:

Select one:

1. the probability of observing a value
2. the probability of observing a value equal to our sampled value
3. the probability of observing a value equal or more extreme than our sampled value
4. the probability of observing a value equal or more extreme than our sampled value, under the null hypothesis ✓

Probability of rejecting a false null hypothesis is:

Select one:

a. the effect size
b. the Type II error
c. the Power ✓
d. the Type I error

A researcher in the field of educational psychology is interested in the effects of 'flipped learning' on intellectual development. Flipped learning is less structured than a traditional lecture, and she believes that this may improve problem-solving skills. A department in her university switched all teaching to a flipped design format last year, and she is interested to see how this has impacted on the students problem-solving skills.

She obtains a random sample of 49 postgraduate students and administers a conceptual problem-solving test to them. It is known that post graduate students in traditional lecture have a mean score of 82 on the test. The students in this study score a mean of 86 with a standard deviation of 5. Compute the 95% CI under the null hypothesis to be able to choose the correct statement below.

Select one:

a. Do not reject the null hypothesis and claim that flipped sessions affect problem-solving skills.
b. Reject the null hypothesis and claim that flipped sessions affect problem-solving skills. ✓
c. There wasn't enough statistical power for you to find an effect.

W4

A researcher wants to study the difference in 'weight' (in kg) across males and females (sex assigned at birth SAAB). Which is th appropriate test to use?

Select one:

- a. One sample t-test
- b. Independent samples t-test ✓
- c. McNemar test
- d. Paired samples t-test

Based on the output, the correct interpretation for the difference in 'weight after' across males and females, is:

		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
Weight after	Equal variances assumed	1.04	.310	.417	298	.677	.09537	.22886	-.3550	.54575
	Equal variances not assumed			.418	297.9	.676	.09537	.22805	-.3534	.54416

Select one:

- a. Non significant (F=1.04, df=298, p=0.677)
- b. Non significant (F=1.04, df=297.9, p=0.676)
- c. Non significant (t=0.418, df=297.9, p=0.676)
- d. Non significant (t=0.417, df=298, p=0.677) ✓

Based on the table, which percentages should be compared?

		SAAB				Total	
		Female		Male			
		Count	% within SAAB	Count	% within SAAB	Count	% within SAAB
Exercised before	No	119	83.2%	103	65.6%	222	74.0%
	Yes	24	16.8%	54	34.4%	78	26.0%
Total		143	100.0%	157	100.0%	300	100.0%

Select one or more:

- a. 83.2% versus 65.6% ✓
- b. 83.2% versus 16.8%
- c. 65.6% versus 34.4%
- d. 26% versus 100%

Based on the output for the association between exercise 'before' and 'after', the correct interpretation is:

		Exercised after				Total	
		No		Yes			
		Count	% of Total	Count	% of Total	Count	% of Total
Exercised before	No	119	39.7%	103	34.3%	222	74.0%
	Yes	48	16.0%	30	10.0%	78	26.0%
Total		167	55.7%	133	44.3%	300	100.0%

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.473[a]	1	.225		
Continuity Correction[b]	1.169	1	.280		
Likelihood Ratio	1.484	1	.223		
Fisher's Exact Test				.236	.140
Linear-by-Linear Association	1.468	1	.226		
McNemar Test				.000[c]	
N of Valid Cases	300				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 34.58.
b. Computed only for a 2x2 table
c. Binomial distribution used.

Select one:

a. The percentage of those exercising 'before' is not different than that of those exercising 'after' (39.7% vs 16% , Pearson Chi Square=1.473, df=1, p=0.225) ×

b. The percentage of those exercising 'before' is different than that of those exercising 'after' (26% vs 44.3% , McNemar p<0.001)

c. The percentage of those exercising 'before' is not different than that of those exercising 'after' (16.0% vs 34.3% , Fisher's exact p=0.236)

d. The percentage of those exercising 'before' is different than that of those exercising 'after' (26% vs 44.3% , Pearson Chi Square=1.473, df=1, p=0.225)

Crosstabulation

		SAAB				Total	
		Female		Male			
		Count	% within Ethnicity	Count	% within Ethnicity	Count	% within Ethnicity
Ethnicity	White	53	42.7%	71	57.3%	124	100.0%
	Black	44	49.4%	45	50.6%	89	100.0%
	Asian	46	58.2%	33	41.8%	79	100.0%
	Other	0	0.0%	8	100.0%	8	100.0%
Total		143	47.7%	157	52.3%	300	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	12.136[a]	3	.007
Likelihood Ratio	15.219	3	.002
Linear-by-Linear Association	.801	1	.371
N of Valid Cases	300		

a. 2 cells (25.0%) have expected count less than 5. The minimum expected count is 3.81.

one:

The ethnicity with the highest percentage of women was Asian (58.2%; Pearson chi-square=12.136, df=3, p=0.007)

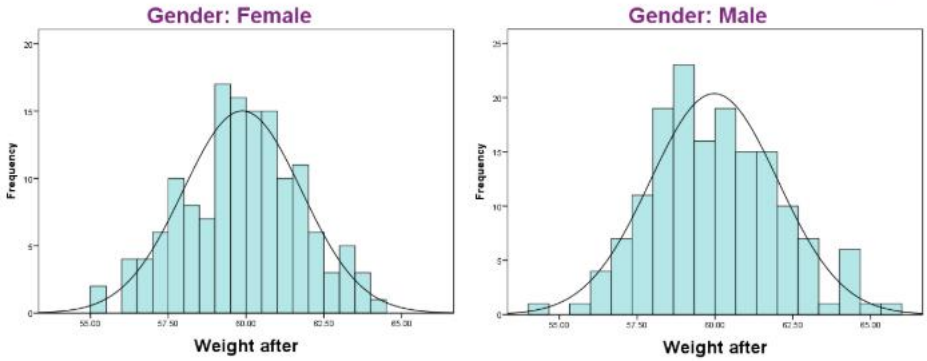
All of the above

None of the above ✓

Among Asian people, the percentage of women was significantly higher than men (58.2% vs 41.8%, Pearson chi-square=12.136, df=3, p=0.007)

W5

To study the differences between 'males' and 'females' in 'weight after', the appropriate test is



Select one:

- a. Independent samples t-test ✓
- b. Paired samples Wilcoxon sign rank test
- c. Independent samples Mann-Whitney test
- d. Paired samples t-test

To study the differences in the' alcohol consumption', 'before' and 'after' the programme, the appropriate test is

More than 2 units of alcohol on a weekend

		After						Total	
		Never		Sometimes		Always			
		Count	% within Before	Count	% within Before	Count	% within Before	Count	% within Before
Before	Never	65	95.6%	3	4.4%	0	0.0%	68	100.0%
	Sometimes	9	5.7%	148	94.3%	0	0.0%	157	100.0%
	Always	0	0.0%	13	17.3%	62	82.7%	75	100.0%
Total		74	24.7%	164	54.7%	62	20.7%	300	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	461.6[a]	4	.000	.000		
Likelihood Ratio	438.05	4	.000	.000		
Fisher's Exact Test	421.74			.000		
Linear-by-Linear Association	250.2[b]	1	.000	.000	.000	.000
McNemar-Bowker Test	16.000	2	.000			
N of Valid Cases	300					

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 14.05.

b. The standardized statistic is 15.819.

Select one:

- a. McNemar-Bowker test ✓

Based on the table below, the correct comparison of percentages is:

Ethnicity * Exercised after Crosstabulation

		Exercised after				Total	
		No		Yes			
		Count	% within Exercised after	Count	% within Exercised after	Count	% within Exercised after
Ethnicity	White	73	43.7%	51	38.3%	124	41.3%
	Black	48	28.7%	41	30.8%	89	29.7%
	Asian	43	25.7%	36	27.1%	79	26.3%
	Other	3	1.8%	5	3.8%	8	2.7%
Total		167	100.0%	133	100.0%	300	100.0%

Select one:

- a. 43.7% vs 38.3% ✓
- b. 43.7 vs 1.8%
- c. 41.3% vs 100%
- d. 3.8% vs 2.7%

To study the association between 'ethnicity' and 'exercise after', the appropriate test is

Ethnicity * Exercised after Crosstabulation

		Exercised after				Total	
		No		Yes			
		Count	% within Exercised after	Count	% within Exercised after	Count	% within Exercised after
Ethnicity	White	73	43.7%	51	38.3%	124	41.3%
	Black	48	28.7%	41	30.8%	89	29.7%
	Asian	43	25.7%	36	27.1%	79	26.3%
	Other	3	1.8%	5	3.8%	8	2.7%
Total		167	100.0%	133	100.0%	300	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	1.743[a]	3	.627	.628		
Likelihood Ratio	1.740	3	.628	.633		
Fisher's Exact Test	1.763			.621		
Linear-by-Linear Association	1.074[c]	1	.300	.323	.166	.031
McNemar-Bowker Test	.	.	.[b]			
N of Valid Cases	300					

a. 2 cells (25.0%) have expected count less than 5. The minimum expected count is 3.55.
b. Computed only for a PxP table, where P must be greater than 1.

Select one:

- a. Pearson's chi-square
- b. McNemar-Bowker test
- c. Linear by Linear Association
- d. Fisher's exact test ✓

W6

histogram | scatter plot | bar chart | pie chart | box plot

A scatter plot ✓ is the appropriate plot to visualise the relationship between two continuous variables.

The Pearson coefficient is the appropriate measure to estimate the linear correlation between any two continuous variables.

Select one:

- True
- False ✓

A simple linear regression model is useful for:

Select one:

- a. Estimating the association between a continuous outcome and a continous explanatory variable. ✓
- b. Predicting a value of an explanatory variable, given a value of the dependent variable.
- c. Predicting a value for the independent variable, given a value for the dependent variable.
- d. Predicting the outcome of any dependent variable with continuous predictor variables.

The coefficients of the least squares regression line are determined by minimising the sum of the squares of the:

Select one:

- a. y-coordinates
- b. Residuals ✓
- c. Differences
- d. x-coordinates

The equation of the regression line is Y= 85 + (-5)*X. Predict y when x = 5.

Answer: 60 ✓

W7

What is multiple linear regression?

Select one:

a. Method of studying the relationship between one dependent and one independent variable.

b. Method for studying the relationship between two or more dependent variables and one independent variable, simultaneously.

c. Method of studying the relationship between one dependent variable and two or more independent variables, simultaneously. ✓

d. Method of studying the relationship between two numerical continuous outcomes.

What is a confounder?

Select one:

a. A confounder is any other variable that does not have an effect on your dependent variable.

b. A confounder is a third variable that can influence both the independent variable of interest (exposure) and the dependent variable (outcome). ✓

c. A confounder is an extra independent variable.

d. A confounder is a third variable on the pathway from the independent variable interest (exposure) to the dependent variable (outcome).

R^2 is a coefficient for what?

Select one:

a. Measuring how well the regression line/hyperplane approximates the real data points. ✓

b. Checking that the most appropriate set of independent variables has been chosen.

c. Model Selection.

d. Measuring the validity of the model in simple and multiple linear regression.

The model $y = 3 + 5x - 2z + \epsilon$ indicates?

Select one:

- a. A change of 5 units in x leads to a 2 unit decrease in y.
- b. x and z are significant predictors of y.
- c. x, y and z are significantly associated.
- d. For every one unit increase in *x*, the predicted value for *y* increases by 5 when *z* is held constant. ✓

Model A showed an Adjusted R^2 of 0.58. Model B showed an Adjusted R^2 of 0.78

Select one:

- a. Both models are unsuitable.
- b. The models are similar.
- c. Model B is better. ✓
- d. Model A is better.

W8

A mediator

Select one:

- a. has an effect on both the dependent Y variable & independent X1 variable.
- b. influences the independent variable.
- c. is caused by the dependent variable.
- d. explains a portion of the total association between Y and X1. ✓

Under a mediated model

Select one:

- a. A simple linear regression model can denote path b.
- b. The total effect of the independent variable X1 on the dependent variable Y is denoted by path c'.
- c. The causal effect can be split into an indirect and direct effects. ✓
- d. The total effect of the independent variable X1 on the dependant variable Y is denoted by the path a.

Complete Mediation is determined when

Select one:

- a. When the path c' is significantly different from 0.
- b. When the path c' is not significantly different from 0. ✓
- c. When the path c is significantly different from 0.
- d. When the path c is not significantly different from 0.

Which Baron and Kenny steps are essential to establish mediation?

Select one:

a. All steps.

b. Steps 2 and 3. ✓

c. Steps 1 and 4.

d. Steps 1 and 3.

The Sobel test is used

Select one:

a. as a method to show complete mediation.

b. as a method for testing *a* and *b* separately.

c. as a method for testing the direct effect.

d. as a method for testing the indirect effect. ✓

Mediation has occurred when:

Select one:

1. The strength of the relationship between the predictor and the outcome is reduced by exactly half when the mediator is included in the model.

2. The relationship between the predictor and the outcome remains the same when the mediator is included in the model.

3. The relationship between the predictor and the outcome is completely wiped out when the mediator is included in the model.

W9

Which of the following examples is a case of effect modification?

Select one:

1. If a man and a woman have the same amount of water per week, the effect on their weight is different. ✓
2. A psychological treatment reduces anxiety and this has en effect on social interaction.
3. A training on driving increases the chances of a person aged 60 or older to renew their license.

To test moderation:

Select one:

1. The 4 steps of Kenny&Baron need to be checked.
2. A new variable needs to be considered. This new term is the cross-product between X1 and the modifier Z . It is denoted as X1 x Z.
3. A simple linear regression model is fitted.

Given the following model: $Y=B_0+B_1X_1+B_2Z+B_3X_1 \times Z + E$

Select one:

1. B_1 is interpreted as the effect of X_1 on Y.
2. B_1 and B_2 are no longer useful unless zero values of the respective predictors are of particular interest. ✓
3. B_3 is interpreted as the difference of the effect of X_1 by levels of Y.

Given the model: years_married= 6 - 2.4 problems_inlaw_family -3.1 cheating_with_others -3.7 problems_inlaw_family * cheating_with_others

Select one:

1. The effect of problems_inlaw_family on years_married is $\beta_1 = -2.4$
2. The effect of problems_inlaw_family on years_married is $= \beta_1 + (\beta_3 *$ cheating_with_others$) = -2.4 +$ (-3.7*cheating_with_others)
3. The effect of problems_inlaw_family on years_married is = 6 -2.4 -3.1+(-3.7*cheating_with_others)

W10

Binary Logistic regression is used when you want to:

- a. Predict a dichotomous variable from continuous or categorical variables. ✓
- b. Predict a continuous variable from dichotomous variables.
- c. Predict any categorical variable from several other categorical variables
- d. Predict a continuous variable from dichotomous or continuous variables

The odds ratio in Binary logistic regression is:

- a. The ratio of the probability of an event not happening to the probability of the event happening.
- b. The ratio of the probability of an event happening to the probability of the event not happening. ✗
- c. The probability of an event occurring.
- d. The ratio of the odds after a unit change in the predictor variable

Which of the following methods do we use to best fit the data in Logistic Regression?

- a. Chi squared Statistic
- b. Pseudo R squared
- c. Ordinary Least Squared Error
- d. Maximum Likelihood Estimation ✓

In a study to determine whether anxiety is associated with subsequent development of depression, the estimated relative risk for those with prior anxiety compared to those who never had anxiety was found to be 1.9. From this we can conclude:

- a. The depression rate is higher among patients who have had prior anxiety.
- b. Those with prior anxiety have a lower risk of developing depression than those who did not have prior anxiety.
- c. The anxiety rate is higher among patients who have had depression.
- d. Those with prior anxiety have a higher risk of developing depression than those who did not have prior anxiety. ✓

There were 5842 men and women surveyed regarding whether or not they experience dizziness and if they used anti-depression medication more than twice in the past two weeks. The data are presented below. Compare the effects of using anti depression medication against not using anti-depression medication on dizziness using the odds ratio.

	Anti-depression medication used	Anti-depression medication not used	Total
Dizziness	443	774	1217
No Dizziness	1530	3095	4625
Total	1973	3869	5842

a. Odds of having dizziness when anti depression medication is taken is about 0.89 times larger than the odds when anti depression

b. Odds of having dizziness when anti depression medication is taken is about 1.16 times larger than the odds when anti depression medication is not taken ✓

c. Odds of having dizziness when anti depression medication is taken is about 1.12 times larger than the odds when anti depression medication is not taken

d. Odds of having dizziness when anti depression medication is taken is about 0.86 times larger than the odds when anti depression medication is not taken