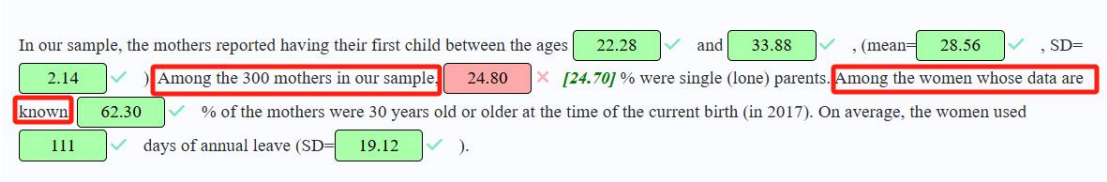
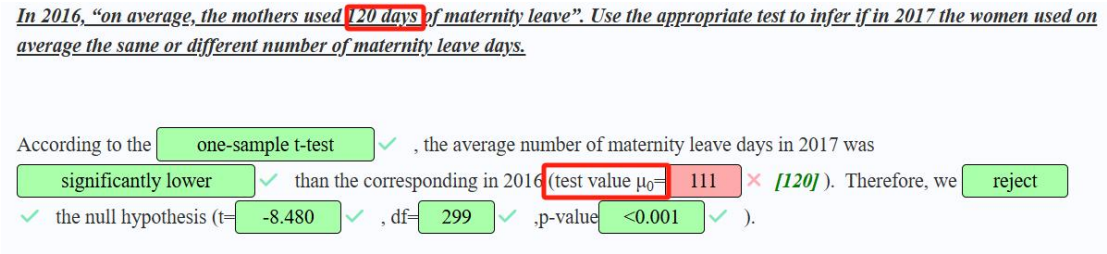
- **数据清洗&描述性统计问题：**

即使在数据清理后，Among the 300 mothers in our sample 当关注“**所有样本**”时，表示看**数据清洗前**的总人数，即看 **percentage 列**；而“Among the women whose data are known” 这句话表示分析的对象是数据**已知群体**，在问题关注**“有效数据”**时，故看 **valid percentage** 列；people in the sample 表示“所有人”，看 **percentage 列**；the people who responded 表有效值，看 **valid percentage** 列



➢ 这里的“**test value**”指μ_0而非样本均值！！



W3

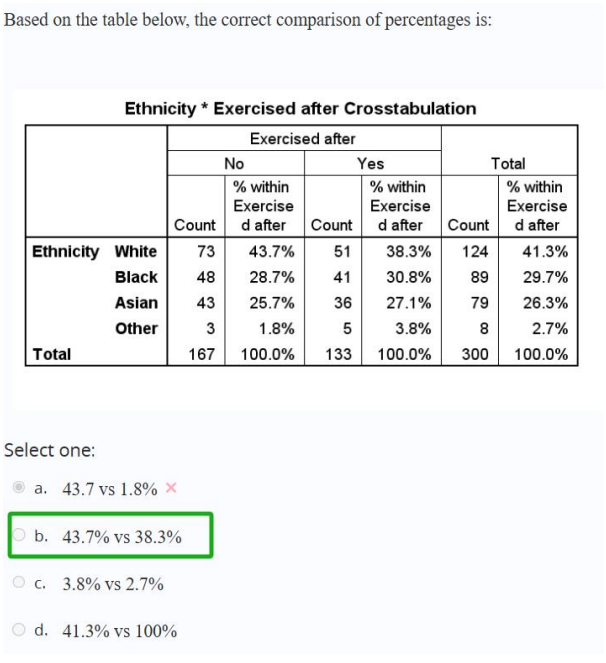
零假设下的 95%置信区间

抽样分布是以μ_0为对称轴；然而**零假设下的 95%置信区间**，仍然以$\bar{x}$为对称轴！**[$\bar{x}$-1.96SE，$\bar{x}$+1.96SE]**

W4

- **卡方组间**应该**按行比较**还是**按列比较**

如何判断：①所有与 **Total** 有关的值**不**参与比较②如果**列相加=100%**，则比较**同一行**的两个值；如果行相加=100%，则比较同一列的两个值



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 1.8% ×

b. 43.7% vs 38.3%

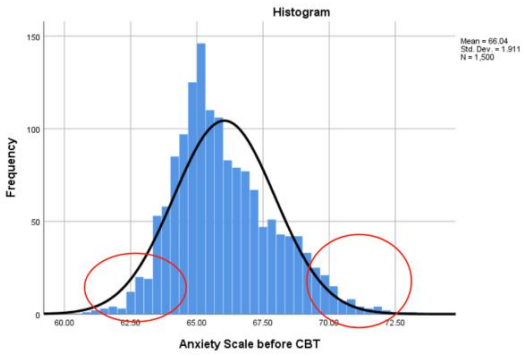
c. 3.8% vs 2.7%

d. 41.3% vs 100%

- **判断是否正态：**

该图片在答案中作为**正态**判断。

注意**左右两边**与正态曲线贴合度；以及 N=1500 是很大的**样本量**



- **参数&非参数表格解读题：**首先判断检验类型，**“前”比“后”**为**配对**，又因比较的是比例，所以用卡方配对 McNemar χ2-test（配对/相关样本）

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)

W6

- **简单线性回归：**通过自变量预测因变量，故 BC 错；D：“any dependent variable”错，回归需要时**连续变量**；A **估计两个连续变量之间关系**正确

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. ×

- 当已知回归方程，给出 x 和 se，**求 CI**，即通过代入 x 求得的预测值，±1.96se。

3. How many health problems will a participant be predicted to have if the number of cigarettes they smoke is 10 and 30. Calculate a confidence interval for the prediction given the s.e. is 0.435

10 cigarettes will lead to 3.109 + (10 x 1.578) = 18.89 health problems
95% CI (18.89 ± 1.96x0.435) = (18.04,19.74)

30 cigarettes will lead to 3.109 + (30 x 1.578) = 50.45 health problems
95% CI (50.45 ± 1.96x0.435) = (49.60,51.30)

- 相关结果能否**推断总体** Can this value be inferred to the whole population?

当相关显著时，因为拒绝零假设，所以可以推断总体中也显著相关；而不显著时，因为“不拒绝零假设”，只能说现有数据**无法支持**相关关系的存在

W8

- 首先读题，研究的是中介效应发生的合理情形，是**充分非必要条件。**在存在中介效应时，自变量与因变量之间的关系，会**有一部分由间接效应解释**；且影响路径也会发生改变，故 B 错误；c 选项中，**效应完全消失**，即**完全中介**的情况，自变量仅通过中介变量影响因变量，符合题意。

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.

- 首先，**product 表乘积**；其次，即使在检验得出直接效应不显著区别于 0 的前提下，**总效应 c** 依然按照简单线性回归的结果写。

c | ac | -0.002 | a | -0.14 | subtraction | addition | -0.135 | -0.280 | product | 0.007 | -0.140 | 0.001 | ab | -0.005 | -0.039 | bc' | c' | 0.482 | ac' | 0.067 | b | bc | -0.038

Estimate the indirect effect from the regression outputs.

Indirect effect (ab ✓) can be calculated as the addition × *[product]* of paths a ✓ and b ✓ , which gives:

ab ✓ = (0.482 ✓) x (-0.280 ✓) = -0.135 ✓ .

The total effect c ✓ is therefore -0.135 × *[-0.140]*

W10

二元回归中的 **odd ratio**，**注意与 odd 区分**。B 选项为 odd，而正确的 D 选项，变量发生 1 单位变化后的优势几率，即为 **Exp（B）**，是正确答案。

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