1. Which scientific method is least likely to have confounds in the results?
A) Observational studies
B) Experiments@
C) Voluntary sample survey studies
D) Confounds will not meaningfully differ between the three study types

2. A study is designed to determine whether grades in a mathematics course could be improved by offering special review material. The 120 students enrolled in a large introductory calculus class are also enrolled in one of 10 practical groups. The 10 practical groups are randomly divided into two groups of 5 each. The students in the first set of 5 practical groups are given extra review material during the last 15 minutes of each weekly practical. The students in the remaining 5 practical groups receive the regular lesson material, without the extra review material. The grades of the students who reviewed weekly were higher, on average, than those of the students who did not review every week. What type of study is this?
A) An observational study that is a sample survey
B) An observational study that is not a sample survey
C) An experiment but not a double-blind experiment@
D) A double-blind experiment

3. In order to select a sample of first year students in the Netherlands, a simple random sample of four university cities is selected. From each of these cities, a simple random sample of 20 male and 20 female first year students is selected. The final sample consists of 160 undergraduates. What sampling technique is being used?
A) Simple random sampling
B) Stratified random sampling
C) Multistage sampling@
D) Convenience sampling

4. When possible, what is the best way to establish that an observed association is the result of a cause-and-effect relation?
A) Study the least-squares regression line.
B) Obtain the correlation coefficient, combined with making sure the association is linear.
C) Examine *z*-scores rather than the original variables.
D) Consistency across studies, a theoretical explanation of the underlying cause, and careful examination of possible lurking variables@

5. This year's season of the Great British Bake-Off had four contestants the jury members were particularly impressed by: Ruby, Kim-Joy, Manon, and Rahul. On 9 October 2018, a journalist asks jury members Paul and Prue to secretly rank these four contestants in terms of how good they are. Paul rates the contestants as follows: (1) Rahul; (2) Kim-Joy; (3) Ruby; (4) Manon. Prue rates the contestants in the following order: (1) Kim-Joy; (2) Rahul (3) Manon; (4) Ruby. Please calculate Kendall's tau for the inter-rater agreement of the two jury members.

A) 0
B) 1/3 @
C) 2/3
D) -1/3

Concordant pairs: (Rahul, Ruby), (Rahul, Manon), (Kim-Joy, Ruby), (Kim-Joy, Manon) so #C=4
Discordant pairs: (Rahul, Kim-Joy), (Ruby, Manon) so #D=2
(#C-#D)/(#pairs) = (4-2)/6 = 1/3

6. The number of undergraduates at the University of Groningen is approximately 40,000, while the number at Maastricht University is approximately 5,000. A simple random sample of 50 undergraduates at University of Groningen will be obtained to estimate the proportion of all Groningen students who feel that drinking is a problem among college students. A simple random sample of 50 undergraduates at Maastricht University will be obtained to estimate the proportion of all Maastricht students who feel that drinking is a problem among college students. Assume the sample proportions will be approximately even. What can we conclude about the sampling variability in the sample proportion, $\hat{p}$, calculated from the sample from Groningen as compared to that in the sample proportion from Maastricht?

A) The sample proportion from Groningen will have less sampling variability than that from Maastricht.
B) The sample proportion from Groningen will have more sampling variability than that from Maastricht.
C) The sample proportion from Groningen will have about the same sampling variability as that from Maastricht. @
D) It is impossible to make any statements about the sampling variability of the two samples because the students surveyed were different.

7. In a certain game of chance, your chances of winning are 0.01. Assume outcomes are independent and that you will play the game fifty times. Suppose it costs 1 euro to play the game. If you win, you receive 90 euros (for a net gain of 89 euros). If you lose, you receive nothing (for a net loss of 1 euro). What are your expected winnings after playing fifty games?

A) -50 euros
B) -5 euros@
C) 40 euros
D) 89 euros

(89*0.01-1*0.99) * 50

8. Suppose we throw a fair coin 20 times. What is the probability that we throw 'heads' *more than 10 times* (so $P(X>10)$)?

A) .176
B) .412@
C) .500
D) .588

P(X>10|B(20,0.5)) = P(X=11|B(20,0.5)) + P(X=12|B(20,0.5)) + … + P(X=20|B(20,0.5))

9. A study was done to compare the amount of time per day students at the University of Groningen in their first year (RUG1) and those same students in their third year (RUG3) spend on Instagram. The study was composed of 200 students from RUG that were measured in their first and third year. The time that RUG1 students spent on Instagram had a Normal distribution with a mean of 220 minutes and a standard deviation of 36 minutes. The time that RUG3 students spent on Instagram had a Normal distribution with a mean of 200 minutes and standard deviation of 49 minutes. The correlation between RUG1 and RUG3 students' time on Instagram was 0.7. What is the standard deviation of the difference in Instagram use between students in their first and third year?

A) 13.0
B) 35.0@
C) 49.6
D) 60.8

$Var(RUG1 - RUG3) = Var(RUG1) + Var(RUG3) - 2 \cdot Corr \cdot SD(RUG1) \cdot SD(RUG3) =$
$= 36^2 + 49^2 - 2 \cdot 0.7 \cdot 36 \cdot 49 = 1227.4$; so $SD(RUG1 - RUG3) = \sqrt{1227.4} = 35.0$

10. Tim the bartender knows his craft. He has been trained to tap beers that contain 220ml of beer, and on average, he gets the quantity pretty close to perfect each time. The distribution of beer quantity per glass follows a normal distribution with a mean of 220ml and a standard deviation of 12ml. A group of 10 students walks in, each of them ordering a beer. Tim taps 10 beers. Assume that each of these ten beers follows the same distribution (i.e., there are no order effects in Tim's services). Please give the sampling distribution of the sample mean of the content of these 10 beers?

A) N(220, 1.2)
B) N(220, 3.8)@
C) N(220, 12)
D) N(220, 37.9)

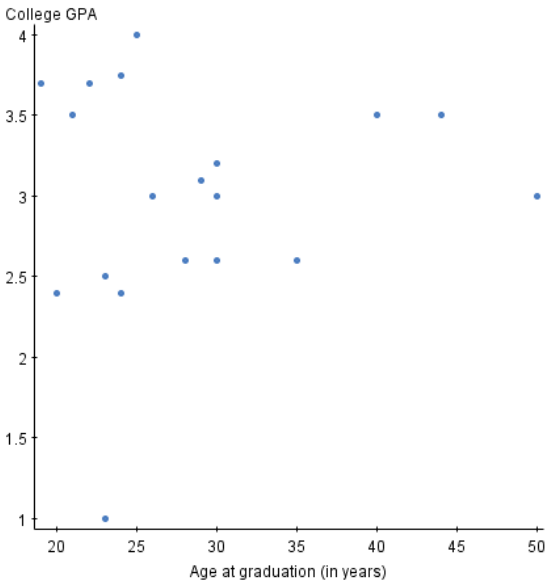
$\sigma_{\bar{x}} = \sigma_x/\sqrt{n} = 12/\sqrt{10} = 3.8$

11. The exam scores for a high school math test are normally distributed with a mean of 6.5 and a standard deviation of 1. Ten students take the math test, what is the probability that the average score of the ten students is higher than a 6?

A) 0.057
B) 0.309
C) 0.691
D) 0.943@

$P(\bar{X} > 6) = P\left(Z > \frac{6-6.5}{1/\sqrt{10}} = -1.58\right) = 1 - 0.057 = 0.943$

12. Data of 20 students is displayed in the scatterplot below.



Which of the statements below concerning this scatterplot is **CORRECT**?

A) The scatterplot shows a strong relationship between both variables
B) If you were to swap the x-axis and the y-axis, the relationship between both variables would be stronger
C) At least one outlier seems to be present in the data set@
D) None of the statements below is correct

13. Malaria is an example of a disease that can be strongly influenced by climate change. In the tables below a summary is provided of a linear regression analysis in which *dew point* (in degrees Celsius) was used to predict the *prevalence* of malaria in West Africa.

	Estimate	**Std. Error**	**T value**	**P value**
Intercept	0.731243	0.092531	7.903	1.47E−13
Dew point	−0.003601	0.001443	−2.495	0.0134

Residual Standard Error:	0.2471
R Squared:	0.0285

The correlation between dew point and prevalence (rounded to 3 decimals) is…

A) -0.004
B) -0.169@
C) 0.029
D) 0.169

Correlation is $-\sqrt{0.0285} = -0.169$ because the regression coefficient of Dewpoint is negative

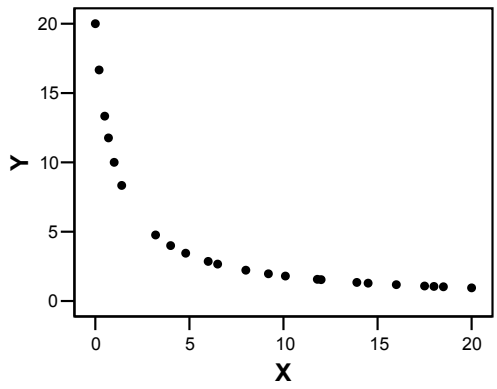
14. Jorge's parents have logged his length in centimeters at various points in time between the age of 36 and 66 months. The table below provides the measurements.

Age (months)	36	48	54	60	66
Length (cm)	86	97	104	110	114

One of the answers below indicates the least squares regression equation. Based on the descriptive data above, select the correct one (hint: drawing a scatterplot might help)

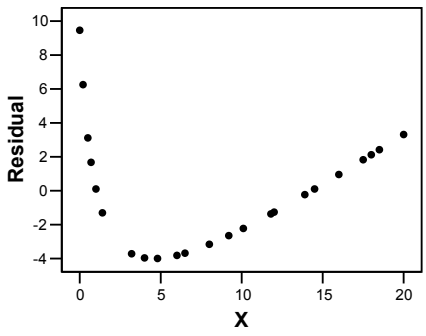
A) Length = 12 ´ (Age)
B) Length = Age/12
C) Length = 60 – 0.56 ´ (Age)
D) Length = 51.5 + 0.96 ´ (Age) @

15. Data for a number of people have been observed on dependent variable *y* and independent variable *x*. A scatterplot of these observations is displayed below

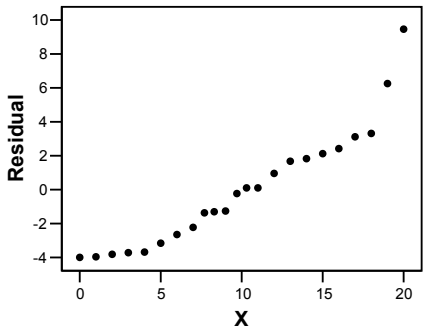


Which of the plots on the next page is the residual plot that belongs to these data with *x* on the *x*-axis (answer options on next page)?

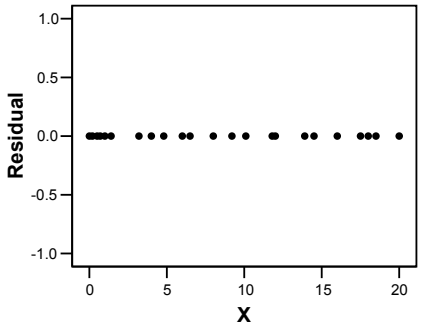
A) @



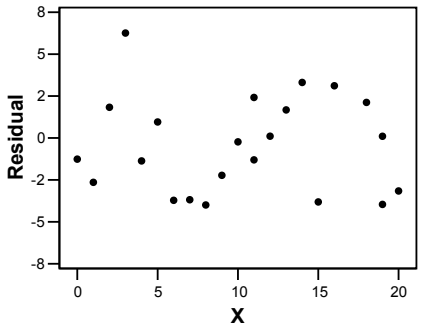
B)



C)



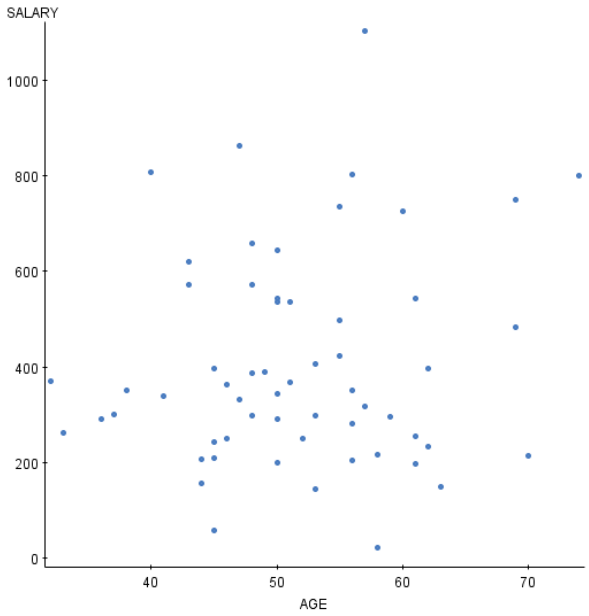
D)



16. Which type of graphical display provides insight into the value of correlation coefficient *r*?

A) histograms
B) boxplots
C) scatterplots@
D) density curves

17. Sixty CEOs, ages varying between 32 and 74 years, have been asked to provide a figure for their yearly salary. Part of the results of a statistical analysis based on this data are provided below.



Simple linear regression results:
Dependent Variable: SALARY
Independent Variable: AGE
SALARY = 242.70212 + 3.1327114 AGE
Sample size: 60
R (correlation coefficient) = 0.1276
R-sq = 0.016270384
Estimate of error standard deviation: 220.64246

Parameter	Estimate	Std. Error	*t*	P-value
Intercept	242.70212	168.7604	1.44	0.1559
Age	3.1327114	3.2264276	0.97	0.3357

Is age a *good* predictor of salary in the sense that it explains a lot of variance in the yearly salary of CEOs?

A) Yes, because the intercept is high
B) Yes, because the correlation is high
C) No, because the intercept is too low
D) No, because R^2 is low@

18. Which statement(s) below concerning probability and randomness is/are **CORRECT**?

A) In statistics the word random can be used to describe regularities that emerge in the long run
B) Probability can be used to describe what happens in the long run
C) With a small number of repetitions, the observed proportion of an outcome can strongly differ from the probability of the outcome
D) All statements above are correct@

19. Which of the two statement(s) below is/are **CORRECT**?

Statement 1: You follow a statistics course. You take the first exam and unfortunately obtain a 1 for a grade. You take remedial teaching and study harder, take the resit and obtain an 8 for a grade. Your scores on the first exam and the resit can be considered independent.
Statement 2: For a given study five statistics courses are required. You can only participate in a statistics course if you successfully completed the previous one. Successfully completing the last statistics course is an independent event.

A) Statement 1 is correct
B) Statement 2 is correct
C) Both statements are correct
D) Neither statement is correct@

20. Say A and B are two independent events with $P(A) = 0.4$ and $P(B) = 0.2$. The probability $P(A^C \text{ and } B^C)$ is equal to…

A) 0.40
B) 0.48@
C) 0.60
D) 0.92

$P(A^c \text{ and } B^c) = P(A^c) \times P(B^c) = 0.6 \times 0.8 = 0.48$

21. The time it take a psychology student to walk from the student house to the Heymans building is approximately normally distributed with mean 20 minutes and standard deviation 3 minutes. What is the probability that a randomly selected student walks from the student house to the Heymans building in a time faster than 15 minutes?

A) 0.0475@
B) 0.2743
C) 0.7257
D) 0.9525

$P(X < 15) = P\left(Z < \frac{15-20}{3} = -1.67\right) = 0.0475$ with table A

22. Variable *X* can take on values 3, 4, 5, and 6 and is distributed according to the probability table displayed below.

Value of *X*	3	4	5	6
Probability	0.10		0.25	0.50

Which of the statements below is/are **CORRECT**?
A) The mean of X is 1.29
B) The standard deviation of X is 1.0275
C) Both A) and B) are correct
D) Neither A) nor B) are correct@

mean $= 3 * 0.1 + 4 * 0.15 + 5 * 0.25 + 6 * 0.5 = 5.15$
variance $= 0.1 * (3 - 5.15)\text{^}2 + \cdots + 0.5 * (6 - 5.15)\text{^}2 = 1.0275$, so stdev $= \sqrt{1.0275}$

23. A company gathered data on 410 possible candidates for an open position. The table below displays the gathered data.

	Minder dan 10 jaar ervaring	**10 of meer jaar ervaring**	**Totaal**
Man	178	112	**290**
Vrouw	99	21	**120**
Totaal	**277**	**133**	**410**

The probability that a male candidate has fewer than 10 years of experience is equal to…
A) 0.430
B) 0.478
C) 0.614@
D) 0.643

24. A statistician records several variables during a festival. One of the variables is age in years (e.g., 19) and another is the educational level (e.g., bachelor degree). Which of the following statements is **CORRECT**?

A) Age and educational level are both measured using ratio scales.
B) Age and educational level are both measured using ordinal scales.
C) Age is measured on a ratio scale and educational level is measured on an ordinal scale.@
D) Age is measured on a ratio scale and educational level is measured on an nominal scale.

25. The five number summary below summarises data from 200 psychology students whose average daily household energy consumption (in kWh) was recorded. Use the 1.5 X IQR rule to determine if there are any outliers present. In this case _____?

Min.	1st Quartile	Median	3rd Quartile	Max.
8.1	10.6	11	11.6	14.4

A) A value of 9.6 would be a potential outlier.
B) Values of 8.7 and 12.4 would both be marked as outliers
C) A value of 13.5 would be marked as an outlier.@
D) None of the above

$IQR = 11.6 - 10.6 = 1, LB = 10.6 - 1.5 = 9.1, UB = 11.6 + 1.5 = 13.1$

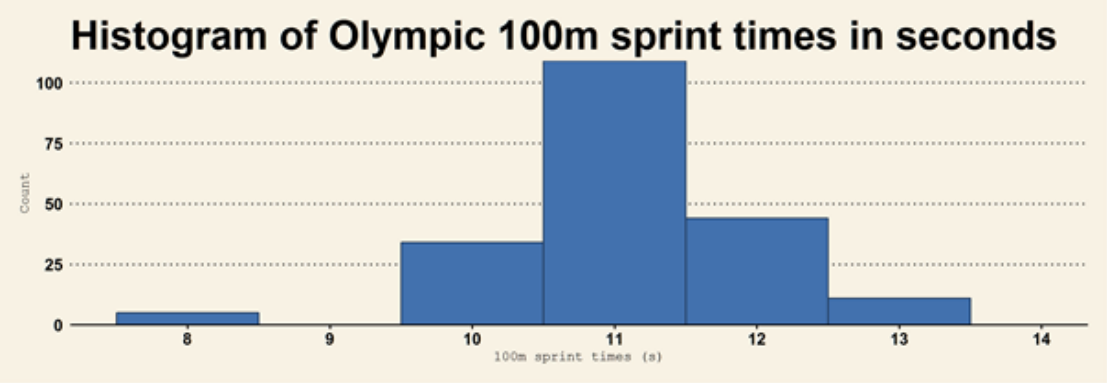
26. The officials at the olympics decided that it would be more interesting to compare the runners' sprint speeds rather than their times. To obtain the speed at which the sprinters' completed the 100m, they apply a transformation on the original data which recorded the sprint times. The transformed variable is given by:

speed = (100 / time taken to sprint 100m) m/s

Which of the following statements is **CORRECT**?

A) The mean of the transformed variable is the same as the mean of the original variable.
B) The transformation is a linear transformation.
C) The median of the transformed variable is the same as the mean of the original variable.
D) The transformation is a non-linear transformation@

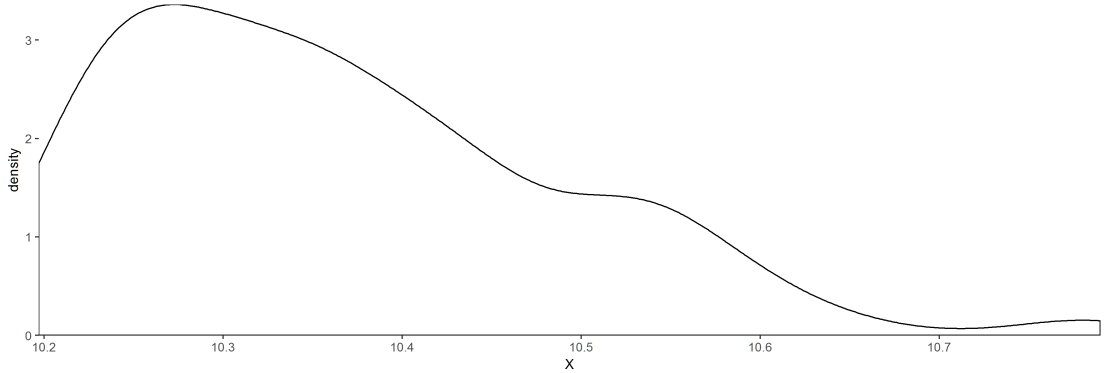
27. The histogram below shows the distribution of data from 200 Olympic 100m sprinters whose sprint times were recorded over the last ten years.



What are some features about the data that you can make out from the histogram?

A) There is a potential outlier in the data@
B) Most sprinters run the 100m in around 11.6 to 12.5s
C) The range of values is between -7.5 and 13.
D) None of the above

28. Examine the density curve below.



Which of the following statements is **CORRECT**?

A) The mean and median are equal
B) The mean is greater than the median@
C) The mean is less than the median.
D) The mean could be either greater than or less than the median

29. The variable Z has a standard Normal distribution. Find the value z such that the event Z>z occurs with a probability of 0.08

A) z = –1.41
B) z = 0.53.
C) z = 0.82.
D) z = 1.41.@

$P(Z > z) = 0.08$, so $P(Z \leq z) = 0.92$ consulting of Table A shows $z = 1.41$

30. The heights of the Maui population is Normally distributed with a mean of 172cm and a standard deviation of 12cm. What percentage of the population is exactly 160cm?

A) 0%@
B) 16%
C) 32%
D) 68%

End of the exam