INKMATHS FLASHCARDS

Statistics essentials

16 cards · fronts and backs

Print double-sided, flipping on the long edge, and each answer lands behind its question. Cut along the card borders.

Definitions in precise wording, the standard results the formulae booklet does not print, a check-yourself question per topic, and the slips that lose marks. The same cards run on the website with spaced-repetition scheduling, at inkphysics.co.uk/maths/flashcards.

STANDARD RESULT

Histograms:

frequency density = ?

STATISTICS ESSENTIALS

STANDARD RESULT

Variance from sums:

σ^2 = ? from Σx and Σx^2

STATISTICS ESSENTIALS

STANDARD RESULT

Outlier fences:

outliers lie beyond ?

STATISTICS ESSENTIALS

STANDARD RESULT

Addition rule:

$P(A \cup B)$ = ?

STATISTICS ESSENTIALS

STANDARD RESULT

Conditional probability:

$P(A|B)$ = ?

STATISTICS ESSENTIALS

STANDARD RESULT

The sample mean's distribution:

mean of n obs of $N(\mu, \sigma^2)$: ?

STATISTICS ESSENTIALS

DEFINITION

Define: Independent events.

STATISTICS ESSENTIALS

STANDARD RESULT

Standardising:

X is $N(\mu, \sigma^2)$: Z = ?

STATISTICS ESSENTIALS

STANDARD RESULT

$\sigma^2 = (\Sigma x^2)/(n) - ((\Sigma x)/(n))^2$

STATISTICS ESSENTIALS

STANDARD RESULT

frequency density = (frequency)/(class width)

STATISTICS ESSENTIALS

STANDARD RESULT

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

STATISTICS ESSENTIALS

STANDARD RESULT

$Q_1 - 1.5 \times IQR$ and $Q_3 + 1.5 \times IQR$

STATISTICS ESSENTIALS

STANDARD RESULT

$N(\mu, \sigma^2/n)$

STATISTICS ESSENTIALS

STANDARD RESULT

$P(A|B) = (P(A \cap B))/(P(B))$

STATISTICS ESSENTIALS

STANDARD RESULT

$Z = (X - \mu)/(\sigma)$, which is $N(0, 1)$

STATISTICS ESSENTIALS

DEFINITION

Events A and B are independent when $P(A \cap B) = P(A) \times P(B)$; equivalently, when $P(A|B) = P(A)$, so B carries no news about A.

STATISTICS ESSENTIALS

DEFINITION

Define: Actual significance level.

STATISTICS ESSENTIALS

RECALL

State the effect of coding on the mean and the standard deviation.

STATISTICS ESSENTIALS

CHECK YOURSELF

A firm has 900 full-time and 300 part-time staff. How many of each go into a stratified sample of 60?

STATISTICS ESSENTIALS

CHECK YOURSELF

$X \sim B(20, 0.15)$. Write $P(X \geq 5)$ in calculator-ready cumulative form.

STATISTICS ESSENTIALS

CHECK YOURSELF

One fair die is rolled; X is the score. Name the distribution of X. Does the total of two dice have the same one?

STATISTICS ESSENTIALS

SPOT THE ERROR

Say what is wrong with this claim:

In a histogram, the tallest bar holds the most values.

STATISTICS ESSENTIALS

SPOT THE ERROR

Say what is wrong with this claim:

A strong correlation shows that one variable causes the other.

STATISTICS ESSENTIALS

SPOT THE ERROR

Say what is wrong with this claim:

Once you have the regression line you can predict y for any x you like.

STATISTICS ESSENTIALS

RECALL

If $y = (x - a)/b$ then the mean of x is b times the mean of y plus a. The standard deviation of x is |b| times the standard deviation of y, since no spread comes out negative and exam codes take b above zero anyway. Subtracting a shifts everything and changes no spread at all, so the a never appears in the second line.

STATISTICS ESSENTIALS

DEFINITION

The probability of landing in the critical region when the null hypothesis is true. Because the binomial is discrete you rarely hit 5 per cent exactly, so the real false-alarm rate usually sits a little under the nominal one.

STATISTICS ESSENTIALS

CHECK YOURSELF

$P(X \geq 5) = 1 - P(X \leq 4)$. At least five means not four or fewer, so the cut is at 4 and not at 5.

STATISTICS ESSENTIALS

CHECK YOURSELF

$60 \times 900/1200 = 45$ full-time and $60 \times 300/1200 = 15$ part-time. Each stratum keeps its share of the population.

STATISTICS ESSENTIALS

SPOT THE ERROR

Height is frequency density, so a tall narrow bar can hold fewer values than a low wide one. Area, height times width, is what counts values.

STATISTICS ESSENTIALS

CHECK YOURSELF

X is discrete uniform: six values, each with probability 1/6, because every score is equally likely. The total of two dice is not — 7 has six ways and 2 has one — so equal likelihood is the test, and naming it is all the specification asks.

STATISTICS ESSENTIALS

SPOT THE ERROR

Inside the range of the data the line is supported by evidence. Outside it you are guessing that a pattern you never observed keeps going. Examiners award the mark for the word extrapolation and for saying the prediction is unreliable, so name the range.

STATISTICS ESSENTIALS

SPOT THE ERROR

Correlation measures how two variables move together, nothing more. A third factor can drive both, the way summer heat lifts ice cream sales and drownings at once. Causation needs an experiment, not an r value.

STATISTICS ESSENTIALS