Excel Cheat Sheet

Descriptive Statistics & Probability for Business · everything that has shown up on past exams

Samkedemia EDUCATION

Functions

IF

Checks if something is true or false, then shows one result if it is and another if it isn't.

`=IF(A1>50;"Pass";"Fail")` If the score in A1 is above 50, shows "Pass". Otherwise "Fail".

`=IF(C2<>$F$7;"negative";"positive")`

`<>` means "is not equal to". If C2 is different from F7, shows "negative". If they match, shows "positive".

AVERAGE and COUNT

AVERAGE gives the average of a group of numbers. COUNT tells you how many numbers there are, so it gives you n.

`=AVERAGE(B2:B101)` Averages every value from B2 down to B101.

`=COUNT(E2:E78)` How many cells in E2:E78 hold a number. Divide a count by this to get a relative frequency.

STDEV.S and STDEV.P

Shows how spread out your numbers are. Low means they're close together, high means they're all over the place.

`=STDEV.S(B2:B101)` For a sample: divides by n − 1.

`=STDEV.P(B2:B101)` For a whole population: divides by n.

On the exam, we normally use .S.

Exam tip: got .P but need .S? Multiply by $\sqrt{\frac{n}{n-1}}$. Example with n = 200: $12.40 \cdot \sqrt{\frac{200}{199}} = 12.43$

COVARIANCE.S and COVARIANCE.P

Checks if two things move together. For example, when it gets hotter, do ice cream sales go up too?

`=COVARIANCE.S(C2:C101;D2:D101)` For a sample: divides by n − 1.

`=COVARIANCE.P(C2:C101;D2:D101)` For a whole population: divides by n.

On the exam, we normally use .S.

COUNTIF

Counts the cells that match one condition.

`=COUNTIF(A1:A5;"Apple")` Counts how many cells in A1:A5 say "Apple".

	A	
1	Apple	✓
2	Pear	
3	Apple	✓
4	Apple	✓
5	Pear	

= 3

Exam tip: cumulative frequency. `=COUNTIF($A$2:$A$1001;"<="&D2)/1000` gives the share of the 1000 values that are at most D2. `&` glues `"<="` to the number in D2.

Exam tip: from cumulative to one bin. Subtract the row above: `=G3-G2`. The first bin is just G2.

COUNTIFS

Like COUNTIF, but with several conditions. A row only counts if it passes all of them.

`=COUNTIFS(A1:A5;"Apple";B1:B5;">100")`

Counts the rows where A says "Apple" and B is over 100.

	A	B	
1	Apple	150	✓
2	Pear	200	
3	Apple	90	
4	Apple	120	✓
5	Pear	50	

= 2 Row 3 says "Apple", but 90 is not over 100.

Exam tip: count one bin. Count every value greater than F2 and up to G2: `=COUNTIFS($C$2:$C$101;">"&F2;$C$2:$C$101;"<="&G2)`. That count is the absolute frequency. Divide by n for the relative frequency.

Exam tip: a whole cross table. One COUNTIFS formula can fill a whole cross table. See the card "Cross table with COUNTIFS".

SUMPRODUCT

Multiplies matching pairs from two ranges, then adds everything up.

`=SUMPRODUCT(B2:B4;C2:C4)` Quantity × price for each item, added up: the total bill.

	B	C
1	Quantity	Price
2	2	5
3	1	10
4	3	4

= 32 $2 \cdot 5 + 1 \cdot 10 + 3 \cdot 4 = 32$

Exam tip: mean of binned data. Relative frequencies in F, bin midpoints in G: `=SUMPRODUCT(F2:F7;G2:G7)`. Each frequency × its midpoint, added up, is the mean.

Exam tip: variance of binned data. $s^2 = \frac{n}{n-1}\left(\sum m_i^2 f_i - \bar{x}^2\right)$, with midpoints m_i and relative frequencies f_i. With the mean in C7: `=450/449*(SUMPRODUCT(B2:B6;B2:B6;C2:C6)-C7*C7)`. Take the square root for the standard deviation.

Exam tip: counts instead of relative frequencies. Divide by n at the end: `=SUMPRODUCT(H2:H5;H2:H5;I2:I5)/232`.

The $ sign and copying

The $ sign

When you copy a formula, Excel moves every cell reference along with it. A `$` freezes the part right after it, so that part stays.

	COPY 1 ROW DOWN	COPY 1 COLUMN RIGHT
`C2`	2 → 3	C → D
`$C2`	2 → 3	**stays C**
`C$2`	**stays 2**	C → D
`$C$2`	**stays 2**	**stays C**

Copying further: count the rows down and the columns right. Every part without a $ moves by exactly that much. 3 rows down and 2 columns right: `C2` → `E5`, `$C2` → `$C5`, `C$2` → `E$2`, `$C$2` → `$C$2`.

Exam tip: make a formula safe to copy down. Freeze the data range: `=COUNTIFS(C2:C101;F2)` becomes `=COUNTIFS($C$2:$C$101;F2)`.

Exam tip: a range that grows. `=STDEV.S($A$2:A4)` copied 1 row down becomes `=STDEV.S($A$2:A5)`.

Cross table with COUNTIFS

One COUNTIFS formula can fill a whole cross table, if the $ signs are in the right places.

`$B$2:$B$93`	The data never moves: freeze both.
`$G4`	Always read the row labels in column G.
`H$3`	Always read the column labels in row 3.

```
=COUNTIFS($B$2:$B$93;$G4;$D$2:$D$93;H$3)
```

Counts the rows where column B matches the row label in G and column D matches the column label in row 3.

	G	H	I
3		18	23
4	32	**10**	1
5	39	10	5

`= 10` 10 girls have weight bin 32 and fat bin 18. Copy H4 right and down to fill the table.

Exam tip: add or remove $ signs. Copying right and down: use the pattern above. Copying only down: only row numbers move, so a $ before a letter does nothing and `C$2:C$78` is enough. Copying only right: the other way around.

Exam tip: labels that are bin bounds. Compare with the lower bound (H) and the upper bound (I): `=COUNTIFS($C$2:$C$78;">="&$H34;$C$2:$C$78;"<"&$I34)` .

Exam tip: relative frequencies. Divide by the number of observations: add `/COUNT($B$2:$B$93)` at the end.

Data Analysis Toolpak

Descriptive Statistics

Data Analysis Toolpak → Descriptive Statistics → tick Summary statistics.

Mean	The average.
Median	The middle value.
Mode	The most frequent value. `#N/A` : no value repeats.
Standard Deviation	Sample standard deviation, the same as STDEV.S.
Sample Variance	Standard deviation squared.
Range	Maximum − Minimum.
Minimum	The smallest value.
Maximum	The largest value.
Count	The number of observations, n.

Exam tip: mode of binned data. Excel's Mode is the most frequent single value. For binned data the mode is the middle of the highest bin, so the two can differ.

Regression

Data Analysis Toolpak → Regression. Input Y Range: what you predict. Input X Range: what you predict it with. Tick Labels if row 1 holds names.

Multiple R	The correlation without its sign. The sign is the sign of the X Variable coefficient.
R Square	Share of the change in Y explained by X. Higher = better fit.
Observations	Number of rows used, n.
Intercept	a in $\hat{y} = a + bx$: the value of Y when X is 0.
X Variable	b in $\hat{y} = a + bx$: the slope, how much Y changes when X goes up by 1.

Correlation and Covariance tables

Data Analysis Toolpak → Correlation. Shows the correlation of every pair of variables at once.

	A	B	C	D
1		Age	Income	Spending
2	Age	**1**		
3	Income	0.42	**1**	
4	Spending	**0.68**	0.35	**1**

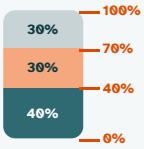
Find where a row and a column meet: Spending and Age gives 0.68. Above the diagonal stays empty, because each pair appears once. The diagonal is always 1.

Exam tip: correlation or covariance table? Data Analysis Toolpak → Covariance makes a table that looks the same. Look at the diagonal: 1s mean correlations, other numbers mean covariances (the diagonal then holds the variances, divided by n).

Charts and pivot tables

100% stacked column chart

Each column is one group and always goes up to 100%. The pieces show how that group splits up.



Where do the pieces meet? Add the shares from the bottom up: 40%, then 40 + 30 = 70%, then 100%.

Pivot table

Excel's quick way to make a cross table: Insert → PivotTable. Drag one variable to Rows, one to Columns, and one to Values (set to Count).

Count of id	Each cell counts the rows with that row label and that column label.
Grand Total	The total of each row and column.
Filter	Hides part of the data. Untick some years and every count is recalculated without them.

	H	I	J	K	L	M	N
26		1	2	3	4	5	Total
27	0	1	3	17	3		**24**
28	1	3	3	7	6	2	**21**
29	**Total**	**4**	**6**	**24**	**9**	**2**	**45**

Parking (rows) by finishing (columns). Divide each count by its row total to get the shares for a 100% stacked column chart.