Bivariate statistics—Things You Should Know

1. Scatter diagrams and recognizing positive/negative linear correlation using them.

2. Pearson's coefficient of correlation r:
- measures linear correlation—doesn't make sense for nonlinear correlation;
- lies in the range $[-1, 1]$;
- is positive for positive linear correlation and negative for negative linear correlation;
- interpretation:
  - $|r| \in [0, 0.25) \implies$ no linear correlation,
  - $|r| \in [0.25, 0.5) \implies$ weak linear correlation,
  - $|r| \in [0.5, 0.75) \implies$ moderate linear correlation,
  - $|r| \in [0.75, 1) \implies$ strong linear correlation,
  - $|r| = 1$ (i.e., $r = -1$ or $r = 1$) $\implies$ perfect linear correlation,
- interpreting r consists in determining the direction (positive or negative) and strength of *linear* correlation;
- calculating r: using the TI-83 (*remember to turn* `DiagnosticOn`*!*) or the formula $r = \frac{s_{XY}}{s_X s_Y}$, where s_X and s_Y stand for the standard deviations of X and Y and s_{XY} for the *covariance* of X and Y (formula for s_{XY} is not on the syllabus).

3. Coefficient of determination r^2—calculation and interpretation.

4. Line of best fit—by eye:
- passes through the point $(\bar{X}, \bar{Y})$, where $\bar{X}$ and $\bar{Y}$ are the means of X and Y respectively;
- needn't pass through any particular point from the data set, but should be close to most or all of them.

5. Line of best fit—using the `LinReg(ax+b)` function on the TI-83.

6. Contingency tables and the χ^2-test:
- the χ^2-test checks whether two features/factors/classifications are *independent*;
- procedure:
  - (i) State the *null hypothesis* H_0: "*foo* and *bar* are independent", where *foo* and *bar* are the factors examined;
  - (ii) fix the so-called *level of significance* $\alpha \in (0, 1)$ (usually given in the task; $\alpha = 0.05$, i.e., 5%, is a typical reasonable value); roughly speaking, it is the margin of error we admit in the statistical inference,
  - (iii) construct a *table of observed values* (might be given),
  - (iv) derive the *table of expected values*,
  - (v) find the U value,
  - (vi) Find the *number of degrees of freedom* ν,
  - (vii) Look up the *critical value* $c_p\chi^2(\nu)$, where $p = 1 - \alpha$ (which means that $p = 0.95$ for $\alpha = 5\%$) and ν is the number of degrees of freedom in the table in the data booklet,

(viii) if $U > c_p\chi^2(\nu)$, *reject* H_0; otherwise *accept* H_0.
Note I: Steps (iv)–(vi) may be performed using the TI-83, but the ability to perform them manually may also be tested!
Note II: Instead of performing steps (vii)–(viii), you may prefer to read the "p-value" from the TI-83 and reject H_0 if this $p < \alpha$, although it's safer to use the standard way. Notice that the TI-83's "p-value" *must not be confused* with the "p-value" from step (vii)!!! Sorry for this mess; remember that Math Studies SL is brutal and full of catches!