

AP Statistics

Module 2 Free Response and Essay Tips

Below you will find a breakdown of different AP topics for this module. The sections include expectations for answering questions over each topic and examples of how these areas should be handled:

Explanatory and Response Variables

	Notes
Explanatory variable (used to explain or predict changes In the values of other variables)	*formerly called the independent variable * plotted along the x-axis when graphing
Response Variable (the variable that measures the outcome)	*formerly called the dependent variable *It is in response to the explanatory variable *plotted along the y-axis when graphing

Interpreting Scatterplots

*we must comment on form, association (or direction), strength, and outliers (if any)

	Notes
Form	* Linear or non-linear
Direction	*points that run from the lower left to upper right has a positive association * points that run from the upper left to the lower right is said to have a negative association
Strength	* Strong – points follow single stream in either direction *Moderate – points follow a stream, but are more scattered *Weak – points scattered all over the place ** Can also use moderately strong or moderately weak to describe strength

AP Statistics

Module 2 Free Response and Essay Tips

Correlation Coefficient

Use your calculator to determine correlation. First, put the data into 2 lists, then use these commands: (TI Calculators: STAT → Calc → LinReg)

	Notes
<p>Correlation coefficient</p> <p>(r)</p>	<p>* $-1 \leq r \leq 1$</p>
<p>Direction</p>	<p>* positive r - positive association between x and y; as x increases, so does y</p> <p>* negative r - negative association between x and y; as x increases, y decreases</p>
<p>Strength</p>	<p>*No set values, but here is a guide (be sure to use absolute value of r):</p> <p>$r < .6 \rightarrow$ weak</p> <p>$.6 < r < .8 \rightarrow$ moderately strong</p> <p>$r > .8 \rightarrow$ strong</p>
<p>Additional points to remember:</p>	<p>* Beware of outliers and influential points when interpreting the correlation coefficient</p> <p>* Correlation is a measure of association, not causation</p>

AP Statistics

Module 2 Free Response and Essay Tips

Least Squares Regression Line

	Notes
<p>Least Squares Regression Line</p> <p>(LSRL)</p>	<ul style="list-style-type: none"> *Used to model data with a linear pattern *Allows us to make predictions and actually use our data *Can be found on the calculator: <p>First, put the data into 2 lists, then use these commands: (TI Calculators: STAT → Calc → LinReg(a+bx))</p> <p>a is y-intercept; b is slope</p> <ul style="list-style-type: none"> *should be written as $\hat{y} = a + bx$ *Variables MUST be defined in context <p>LSRL can also be written in context without x and y in equation:</p> <p>Ex: $\hat{\text{Height}} = 53.24 + 1.65(\text{shoe size})$</p> <p>Once you have the LSRL you can predict what y will be for different values of x. Be sure to review your prediction to ensure it makes sense in the context of the problem.</p>

AP Statistics

Module 2 Free Response and Essay Tips

Residual Plots

	Notes
Residual	<p>*The difference between the response value of the actual data point and predicted value from the LSRL</p> <p style="text-align: center;">Residual = Actual – Predicted</p>
Residual Plot	<p>*called a scattergram</p> <p>*shows relationship between explanatory variable and their respective residual</p> <p>* explanatory is still the x-axis and the residuals now make up the y-axis with center being at zero</p> <p>*should always be created after you have done a regressions</p>
No Pattern to Residual Plot	*LSRL is a good model
Pattern to Residual Plot	*LSRL is NOT a good model; another model would be more appropriate; try exponential or power model
Calculating Residuals with Technology	<p>The residuals are automatically calculated on your calculator when you find the LSRL.</p> <p>Go to your Stat Plot and change the YList to RESID. Leave the XList as your explanatory variable.</p>

AP Statistics

Module 2 Free Response and Essay Tips

Coefficient of Determination

	Notes
<p>Coefficient of Determination</p> <p>(r^2)</p>	<p>*Gives the proportion of variation in y that can be attributed to an approximate linear relationship between x and y</p>
<p>Phrasing that MUST be used in the course and on the AP Exam</p>	<p>*If $r^2 = .####$, then “###.## % of the variation in the y-variable is explained by the LSRL of y on x.”</p> <p>You will replace x and y according to the context of the problem.</p> <p>Ex: $r^2 = .905$</p> <p>“90.5% of the variation in the annual salary is explained by the LSRL of annual salary on years of education”</p>

Nonlinear Data

For the purposes of this course, we will only focus on data that is either linear or can be made linear through simple transformations.

Always check your scatterplot and residual plot in determining the best model to use. Your correlation may be .85 or .9, so you may think you have the right model, but if your residual plot has a pattern, you must transform your data. It may be as simple as taking the square root of one of the variables, or you may have to the logarithm of one or both of variables.

Exponential Model

Once you plot (x,y) and notice a curved pattern in the scatterplot, take the logarithm of your y data, and plot (x,logy). You can do this in a 3rd list in your calculator. If your y data is in List 2, go to List 3, and type the command: log(L2). Now find the LSRL of List 1(or wherever your x variable is) and List 3 (the logarithms of the y)

Your equation will now be: $\log(y) = a + bx$. Remember your Algebra 2 skills to get y by itself. (Step 1: Make both sides an exponent with base 10. Remember $10^{(\log y)}$ simplifies to y.

AP Statistics

Module 2 Free Response and Essay Tips

Power Model

Same process as Exponential Models, but instead of only taking the logarithm of the y-values, you will take the logarithm of both x and y. Again, remember, each time you do a regression, you should follow up that work with a residual plot. This is an **important final check** on your model of the straight-line relationship. If a clear curve appears in the residual plot, then your regression model is not the best model.