

Fox Module 22: INTUITION: NORMAL, POISSON, AND EXPONENTIAL DISTRIBUTIONS

(The attached PDF file has better formatting.)

Fox's textbook teaches you how to use GLMs. This on-line course covers the concepts of GLMs, not the details of forming and running GLMs. You don't have the software needed to run GLMs, unless you use SAS or R (or similar packages).

This posting explains the intuition of normal, Poisson, and exponential distributions. These are conditional distributions of the response variable, conditioned on the fitted value. We show how to determine the GLM values using Excel's *SOLVER* add-in.

Classical linear regression assumes a normal distribution with a constant variance. The distribution of the errors does not depend on the response variable.

GLMs speak of the conditional distribution of the response variable.

- The distribution is conditional on the explanatory variables.
- The variance of the distribution is not constant. It depends on the mean, which is a function of the explanatory variables.

The distribution of errors is replaced by a distribution of observations about their means.

- The fitted line depends on the relation of the variances of observations to their means.
- We use examples of three points fit to a straight line to clarify the intuition.

We examine normal, Poisson, and exponential distributions in this posting.

- The exponential distribution has a variance proportional to the square of the mean.
- The Poisson distribution has a variance proportional to the mean.

THE ILLUSTRATION

Illustration: We fit a linear model to three points: $\{ (1,1), (5,5), \text{ and } (9,9) \}$.

- The points lie on a straight line, and the linear model is $Y = 0 + X + \epsilon$.
- The fit does not depend on the distribution of the residuals. The residuals are zero, and their variance is zero regardless of the conditional distribution of the response variable.

We revise the three points and re-fit the linear model.

- We add 1 point to Y at X = 1, subtract two points at X = 5, and add one point at X = 9.
- The revised three points are $\{ (1,2), (5,3), \text{ and } (9,10) \}$.

Classical regression analysis assumes

- The distribution of the error terms at each point has the same variance.
- The expected (fitted) value of Y does not affect the variance of Y about its mean.

The three points still have a mean of (5,5), but the observed line is bent in the middle.

- The differences from the mean for the original points are $\{ (-4, -4), (0, 0), (4, 4) \}$.
- The differences from the mean for the revised points are $\{ (-4, -3), (0, -2), (4, 5) \}$.

The original line with a constant slope of 1 is now two line segments.

- From (1,2) to (5,3), the slope is 0.25.
- From (5,3) to (9,10), the slope is 1.75.

A linear model means the true slope is constant. The β coefficient is the estimated slope of the full line, so it is a weighted average of 0.25 and 1.75. The conditional distribution of the response variable determines how much weight is given to each line segment.

- For a normal distribution with the same variance at each point, adding 1 point to a Y value of 9 has the same weight as adding 1 point to a Y value of 1.
- If the standard deviation σ of the distribution is constant, the probability of being 1 point away from the mean is the same whether the mean is 1 or 9.

If the true slope is 1, the likelihood of random errors decreasing the first slope by 0.75 is the same as that of random errors increasing the second slope by 0.75.

For other distributions of the response variable, the variance is not the same at $Y=1$ as at $Y=9$, and the weights for the two slopes are not equal. For a Poisson distribution:

- The variance is proportional to its mean and the standard deviation is proportional to the square root of its mean.
- A random error of one point from a Y value of 1 is less likely than a random error of one point from a Y value of 9.

We examine first the likelihood of residuals in a Poisson distribution, and then we show how a Poisson distribution rotates the regression line in this illustration.

Exercise 1.1: Poisson distributions

We examine Poisson distributions with means of 1 vs 9, comparing

- The probability that $Y = \text{the mean } \mu$
- The probability that $Y = \mu + 1$

These are the residuals at fitted line $Y = Z$ in the illustration above at the two end points:

- Fitted $Y = 1$ and observed $Y = 1$ vs 2
- Fitted $Y = 9$ and observed $Y = 9$ vs 10

What is the relation of these two probabilities for each Poisson distribution?

Solution 1.1: We compare

- The likelihood of a Y value of 2 vs 1 when the mean is 1 with
- The likelihood of a Y value of 10 vs 9 when the mean is 9.

<i>Mean</i>	$\mu = 1$		$\mu = 9$	
<i>Observation</i>	$y = 1$	$y = 2$	$y = 9$	$y = 10$
<i>Probability</i>	0.368	0.184	0.132	0.119

- The likelihood of $y=2$ when $\mu=1$ is 50% of the likelihood of $y=1$ when $\mu=1$.
- The likelihood of $y=10$ when $\mu=9$ is 90% of the likelihood of $y=9$ when $\mu=9$.

Take heed: The final exam ask about the relative likelihoods of residuals given conditional distributions of the response variable, similar to the exercise above.

Exercise 1.2: Line of Best fit (Weighting the Residuals)

Three observed points are $\{(1,2), (5,3), \text{ and } (9,10)\}$. We fit straight lines with three GLMs:

- Normal, Poisson, and exponential distributions of the observed Y values.
- Identity link functions \Rightarrow the fitted Y value is a linear function of the X value.

Note: The link function describes how the response variable (the Y values) relate to the explanatory variable (the X values). This exercise uses the identity link function: the fitted Y value is a linear function of the X value. This exercise does not deal with link functions. It shows how the assumed distribution affects the fitted line.

- The distribution gives the variance of the response variable about its mean.
- Assume the normal distribution has the same variance at each point.

The weights for the residuals depend on the variance of the response variable at the point. These weights determine the estimated slopes and intercepts.

- The β coefficient is the slope of the regression line.
- The α coefficient is the Y-intercept of the regression line.

- A. Which distribution gives the most weight to the residual at $X = 1$? At $X = 9$?
- B. Which distribution gives the most weight to the slope from $X = 1$ to $X = 5$? From $X = 5$ to $X = 9$?
- C. Which distribution gives the highest β ? The lowest β ?
- D. Which distribution gives the highest α ? The lowest α ?

Intuition: No straight line passes through all three points.

- A straight line through $(1,2)$ and $(5,3)$ passes through $(9,4)$.
- A straight line through $(9,10)$ and $(5,3)$ passes through $(1,-4)$.

Ordinary least squares minimizes the sum of squared errors. The straight lines in the bullet points above have residuals of zero at two points and a residual of 6 at the third point.

- The sum of squared errors for both lines above is $0^2 + 0^2 + 6^2 = 36$.
- This sum of squared errors gives equal weight to all points. Equal weights assume the variances are the same at all points.

With GLMs, the weights are not equal at all the points since the variances are not equal.

Part A: The weights depend on the *fitted Y values* at each point.

- We *don't know* the fitted Y value until we fit the line.
- The fitted line depends on the distribution of the response variable about its mean.

This exercise focuses on the GLM concepts, not the mathematics. We do not know the exact fitted value at each sample point, so we do not know the exact weights, so we can not minimize the weighted residuals by an algebraic formula. GLMs uses iterative weighted least squares: repeated numerical estimates which converge to the solution.

This exercise, as a proxy, uses the observed Y value at each point. The observed values are spread out and the residuals are small, so the proxy works well.

Take heed: The final exam may ask which point receives the most weight in a GLM.

The final exam does not require you to solve complex GLMs. The exam problems focus on the concepts of GLM analysis, not the numerical solutions to complex equations.

For a Poisson distribution of the response variable:

- The observed values are $Y = 2$ at $X = 1$ and $Y = 10$ at $X = 9$.
- The variances are approximately $\sigma^2 = 2$ at $X = 1$ and $\sigma^2 = 10$ at $X = 9$.
- The standard deviations are approximately $\sigma = 1.414$ at $X = 1$ and $\sigma = 3.162$ at $X = 9$.

The variances are approximations. The fitted Y value is near 2 but not exactly 2, so the variance is near 2, not exactly 2.

With a Poisson distribution of the response variable, the estimate is more precise at $X=1$ than at $X=9$ because the variance is smaller at $X=1$ than at $X=9$.

HETEROSCEDASTIC DATA, STANDARDIZED RESIDUALS, AND WEIGHTED LEAST SQUARES

A GLM with a Poisson distribution is like linear regression with heteroscedastic data.

- We use standardized residuals, or the residuals divided by the standard deviation.
- We give more weight to the residual at $X = 1$ when fitting the regression line.
- We give less weight to the residual at $X = 9$.

For an exponential distribution of the residuals:

- The observed values are $Y = 2$ at $X = 1$ and $Y = 10$ at $X = 9$.
- The variances are approximately $\sigma^2 = 4$ at $X = 1$ and $\sigma^2 = 100$ at $X = 9$.
- The standard deviations are approximately $\sigma = 2$ at $X = 1$ and $\sigma = 10$ at $X = 9$.

A GLM with an exponential distribution has an even greater difference of the variances.

- We give much more weight to the residual at $X=1$ when fitting the regression because the variance is so small at $X=1$ (compared to the variance at other points).
- We give even less weight to the residual at $X = 9$ because its variance is so large.

Of the three distributions of the error term:

- The exponential distribution gives the most weight to the observed value at $X = 1$.
- The normal distribution gives the most weight to the observed value at $X = 9$.

Take heed: The ratio of observed values (10 to 2 at $X=9$ compared to $X=1$) is not the ratio of fitted values. For the exponential distribution:

- The low variance at $X=1$ causes the fitted Y value to be close to $Y=2$.
- The high variance at $X=9$ causes the fitted Y value to be farther from $Y=10$.

Illustration: If the GLM causes the fitted values to be (1, 1.8) and (9, 8.1), the *ratio of fitted values* is $8.1 / 1.8 = 4.500$ to 1.

Parts B and C: The distribution of the residuals changes the slope of the fitted line.

The fitted line for the Poisson and exponential distributions passes

- Closest to the point (1, 2).
- Less closely to the point (5,3).
- Least closely to the point (9, 10).

For the Poisson distribution compared to the normal distribution with constant variance:

- The slope of the fitted line is closer to the slope from (1,2) to (5,3) than to the slope from (5,3) to (9,10).
- We give more weight to the slope 0.25 than to the slope 1.75.
- The fitted β is closer to 0.25 than to 1.75
- \Rightarrow The fitted line is flatter than a 45° diagonal line.

For an exponential distribution compared to the Poisson distribution:

- We give even more weight to the slope 0.25 than to the slope 1.75.
- β is even lower.
- The fitted line is even flatter.

Take heed: Fitting the curve changes the means of the Y values at $X = 1$ vs $X = 9$. See the fitted values at the end of this posting.

Part D: The relative values of the intercept α reflect the relative slopes.

- The normal distribution with constant variance has a slope of 1 and an α of zero.
- The Poisson distribution has a flatter line (lower slope) and a higher α .
- The exponential distribution has the flattest line (lowest slope) and the highest α .

The Poisson distribution rotates the fitted line clockwise, making it flatter. The points lie in the first quadrant, so the clockwise rotation raise the Y-intercept.

Choosing other points has different effects on α .

- If we make all the X vales negative (-1 , -5 , and -9) and keep the same Y values (2, 3, and 10), the fitted line slopes downward and the Y-intercept is on the right side. The Poisson distribution rotates the line counter-clockwise and raises the Y intercept.
- If we change the points to (1, 10), (5, 3), and (9, 2), the fitted line slopes downward and the Y-intercept is on the right side. The Poisson distribution rotates the line counter-clockwise and lowers the Y intercept.
- If we change the points to (-1 , 10), (-5 , 3), and (-9 , 2), the fitted line slopes upward and the Y-intercept is on the right side. The Poisson distribution rotates the line clockwise and lowers the Y intercept.

{The homework assignment for this Module is similar to the exercise below, but it does not require computations. Read this exercise and then complete the homework assignment.}

Exercise 1.3: Line of best fit with a Poisson distribution of the response variable

Three observed points are $\{ (1,2), (5,3), \text{ and } (9,10) \}$.

We fit a straight line $Y = a + b \times X$ with a Poisson distribution of the response variable.

- A. What is the loglikelihood at the three observed points as a function of a and b ?
- B. What is the total loglikelihood?
- C. What two equations do we use to maximize this loglikelihood?
- D. Solve the two equations using Excel's *SOLVER* add-in. Some statistical packages solve GLMs directly, including SAS and R.
- E. What are the residuals and the squared residuals at each of the three points?
- F. Is the sum of squared residuals more or less than with classical regression analysis?
- G. What is the ratio of the fitted values at the points $X = 1$ and 9 ?
- H. What is the ratio of the variances of the error terms at the points $X = 1$ and 9 ?
- I. What is the ratio of the residuals at the points $X = 1$ and 9 ?
- J. Compute the sum of squared residuals divided by their variances. Show that this sum is less than the sum with an ordinary least squares regression line.

Part A: For a Poisson distribution of error terms, the loglikelihood = $y \ln(\mu) - \mu - \ln(y!)$, where y is the observation and μ is the fitted mean.

Each Y value is a linear function of the X value: $Y = a + b \times X = \text{intercept} + \text{slope} \times X$

The loglikelihood of observing each of the three points is

- $(1,2) \rightarrow 2 \ln(a + 1b) - (a + 1b) - \ln(2!)$
- $(5,3) \rightarrow 3 \ln(a + 5b) - (a + 5b) - \ln(3!)$
- $(9,10) \rightarrow 10 \ln(a + 9b) - (a + 9b) - \ln(10!)$

Part B: The loglikelihood of observing all three points is the sum of the loglikelihoods.

$$\begin{aligned} & 2 \ln(a + 1b) - (a + 1b) - \ln(2!) \\ + & 3 \ln(a + 5b) - (a + 5b) - \ln(3!) \\ + & 10 \ln(a + 9b) - (a + 9b) - \ln(10!) \end{aligned}$$

Part C: We set the partial derivatives with respect to a and b equal to zero.

Setting partial derivative with respect to a equal to 0 gives

$$2 / (a + b) + 3 / (a + 5b) + 10 / (a + 9b) - 3 = 0$$

Setting partial derivative with respect to b equal to 0 gives

$$2 \times 1 / (a + b) + 3 \times 5 / (a + 5b) + 10 \times 9 / (a + 9b) - (1 + 5 + 9) = 0$$

Part D: We have two equations in two unknowns. They are not linear equations, so we have no closed form solution.

We use Excel's *SOLVER* add-in to find the values of a and b .

- Type the name *intercept* in Cell A11 and an estimate (such as -1) in Cell B11.

- Type the name *slope* in Cell A12 and an estimate (such as 1) in Cell B12.
- In Cell A13, enter the formula

$$= 2 / (\text{intercept} + \text{slope}) + 3 / (\text{intercept} + 5 * \text{slope}) + 10 / (\text{intercept} + 9 * \text{slope}) - 3$$

- In Cell A14, enter the formula

$$= 2 \times 1 / (\text{intercept} + \text{slope}) + 3 * 5 / (\text{intercept} + 5 * \text{slope}) + 10 * 9 / (\text{intercept} + 9 * \text{slope}) - (1 + 5 + 9)$$

- In Cell A15, enter the formula $= A13^2 + A14^2$

Click on the *TOOLS* menu and choose *SOLVER*.

Set Cell A15 to 0 by choosing values for *intercept*, *slope*.

SOLVER returns the solution: *intercept* = 0.834517 and *slope* = 0.833119.

The slope of the fitted line is 0.833, so it is flatter than the least squares regression line.

Part E: The fitted values are

<i>X</i>	<i>Observed Value</i>	<i>Fitted Value</i>	<i>Residual</i>	<i>Residual Squared</i>
1	2	1.668	0.332	0.110
5	3	5.000	-2.000	4.000
9	10	8.333	1.667	2.780
Total	15	15.001	-0.001	6.891

We show the solution using Excel's *SOLVER* add-in because most candidates use *SOLVER*. The most versatile software package for GLMs is R, which is freely available. With R:

- specify the x values as `xv <- c(1,5,9)`
- specify the y values as `yv <- c(2,3,10)`
- run the GLM as `glm(yv ~ xv)`

The variables names *xv* and *yv* are arbitrary; any names are fine. Specify the probability distribution as family = Poisson and the link function as link = identity.

Part F: With ordinary least squares estimation, the residuals are 1, -2, and 1, and the sum of squared residuals is $1^2 + 2^2 + 1^2 = 6$.

With the Poisson distribution of error terms, the sum of squared residuals is 6.891, which is greater.

Parts G, H, I:

We compare the ratios of fitted values, variances, and residuals at the points *X* = 1 and 9.

- Fitted Values: $8.333 / 1.668 = 4.996 \approx 5$.
- The variances are proportional to the fitted values, so the ratio is also 5.
- Residual: $1.667 / 0.332 = 5.021 \approx 5$.

Part J: We compute the residual divided by the variance of the error term, which equals the fitted value.

<i>X</i>	<i>Observed Value</i>	<i>Fitted Value</i>	<i>Residual</i>	<i>Residual Squared</i>	<i>Rsd / Var</i>
1	2	1.668	0.332	0.110	0.199
5	3	5.000	-2.000	4.000	-0.400
9	10	8.333	1.667	2.780	0.200
Total	15	15.000	0.000	6.891	-0.001

The total (residual / variance) is zero, ignoring the rounding error in the table.

The Poisson distribution of error terms causes the fitted line to rotate about its mean four fifths of the way back to the observed values.

Exercise 1.4: Line of best fit with an exponential distribution of the response variable

Three observed points are $\{ (1,2), (5,3), \text{ and } (9,10) \}$.

We fit a straight line $Y = a + b \times X$ with an exponential distribution of the error terms.

- A. What is the loglikelihood at the three observed points as a function of a and b ?
- B. What is the total loglikelihood?
- C. What two equations do we use to maximize this loglikelihood?
- D. Solve the two equations using Excel's *SOLVER* add-in.
- E. What are the residuals and the squared residuals at each of the three points?
- F. Is the sum of squared residuals more or less than with classical regression analysis?
- G. What is the ratio of the fitted values at the points $X = 1$ and 9 ?
- H. What is the ratio of the variances of the error terms at the points $X = 1$ and 9 ?
- I. What is the ratio of the residuals at the points $X = 1$ and 9 ?
- J. Compute the sum of squared residuals divided by their variances. Show that this sum is less than the sum with an ordinary least squares regression line.

Part A: For an exponential distribution of the error terms, the loglikelihood is $-y/\mu - \ln(\mu)$, where y is the observation and μ is the fitted mean.

Each Y value is a linear function of the X value:

$$Y = a + b \times X = \text{intercept} + \text{slope} \times X$$

The loglikelihood of observing each of the three points is

- $(1,2) \rightarrow -\ln(a + 1b) - 2 / (a + 1b)$
- $(5,3) \rightarrow -\ln(a + 5b) - 3 / (a + 5b)$
- $(9,10) \rightarrow -\ln(a + 9b) - 10 / (a + 9b)$

Part B: The loglikelihood of observing all three points is the sum of the loglikelihoods.

$$\begin{aligned} & -\ln(a + 1b) - 2 / (a + 1b) \\ + & -\ln(a + 5b) - 3 / (a + 5b) \\ + & -\ln(a + 9b) - 10 / (a + 9b) \end{aligned}$$

Part C: We set the partial derivatives with respect to a and b equal to zero.

Setting partial derivative with respect to a equal to 0 gives

$$-1 / (a + b) + -1 / (a + 5b) + -1 / (a + 9b) + 2 / (a + b)^2 + 3 / (a + 5b)^2 + 10 / (a + 9b)^2 = 0$$

Setting partial derivative with respect to b equal to 0 gives

$$-1 / (a + b) + -5 / (a + 5b) + -9 / (a + 9b) + 2 * 1 / (a + b)^2 + 3 * 5 / (a + 5b)^2 + 10 * 9 / (a + 9b)^2 = 0$$

Part D: We have two equations in two unknowns. They are not linear equations, so we have no closed form solution.

We use Excel's *SOLVER* add-in to find the values of a and b .

- Type the name *intercept* in Cell A11 and an estimate (such as -1) in Cell B11.
- Type the name *slope* in Cell A12 and an estimate (such as 1) in Cell B12.
- In Cell A13, enter the formula for the partial derivative with respect to a (*intercept*).

- In Cell A14, enter the formula for the partial derivative with respect to b (slope).
- In Cell A15, enter the formula $= A13^2 + A14^2$
- Click on the *TOOLS* menu and choose *SOLVER*.
- Set Cell A15 to 0 by choosing values for *intercept*, *slope*.
- *SOLVER* returns the solution: intercept = 1.137295 and slope = 0.728246.

The slope of the fitted line is 0.728, so it is flatter than the fitted line with a Poisson distribution of error terms.

Part E: The fitted values are

X	<i>Observed Value</i>	<i>Fitted Value</i>	<i>Residual</i>	<i>Residual Squared</i>
1	2	1.866	0.134	0.018
5	3	4.779	-1.779	3.163
9	10	7.692	2.308	5.329
Total	15	14.336	0.664	8.510

We compare the ratios of fitted values, variances, and residuals at the points $X = 1$ and 9.

- Fitted Values: $7.692 / 1.866 = 4.122$.
- Variances are proportional to the squares of the fitted values: $4.122^2 = 16.991 \approx 17$
- Residuals: $2.308 / 0.134 = 17.224 \approx 17$.

We used an exponential distribution in this exercise to keep the mathematics simple. GLM software uses a Gamma distribution. (The exponential distribution is the one-parameter version of the Gamma distribution.) The Gamma distribution is the exponential family proxy for distributions whose variance is proportional to the square of the mean.

Part F: For the three distributions of the error term:

- Normal with constant variance: the sum of squared residuals is $1^2 + 2^2 + 1^2 = 6$.
- Poisson: the sum of squared residuals is 6.891.
- Exponential: the sum of squared residuals is 8.510

Parts G, H, I: We compare the ratios of fitted values, variances, and residuals at the points $X = 1$ and 9.

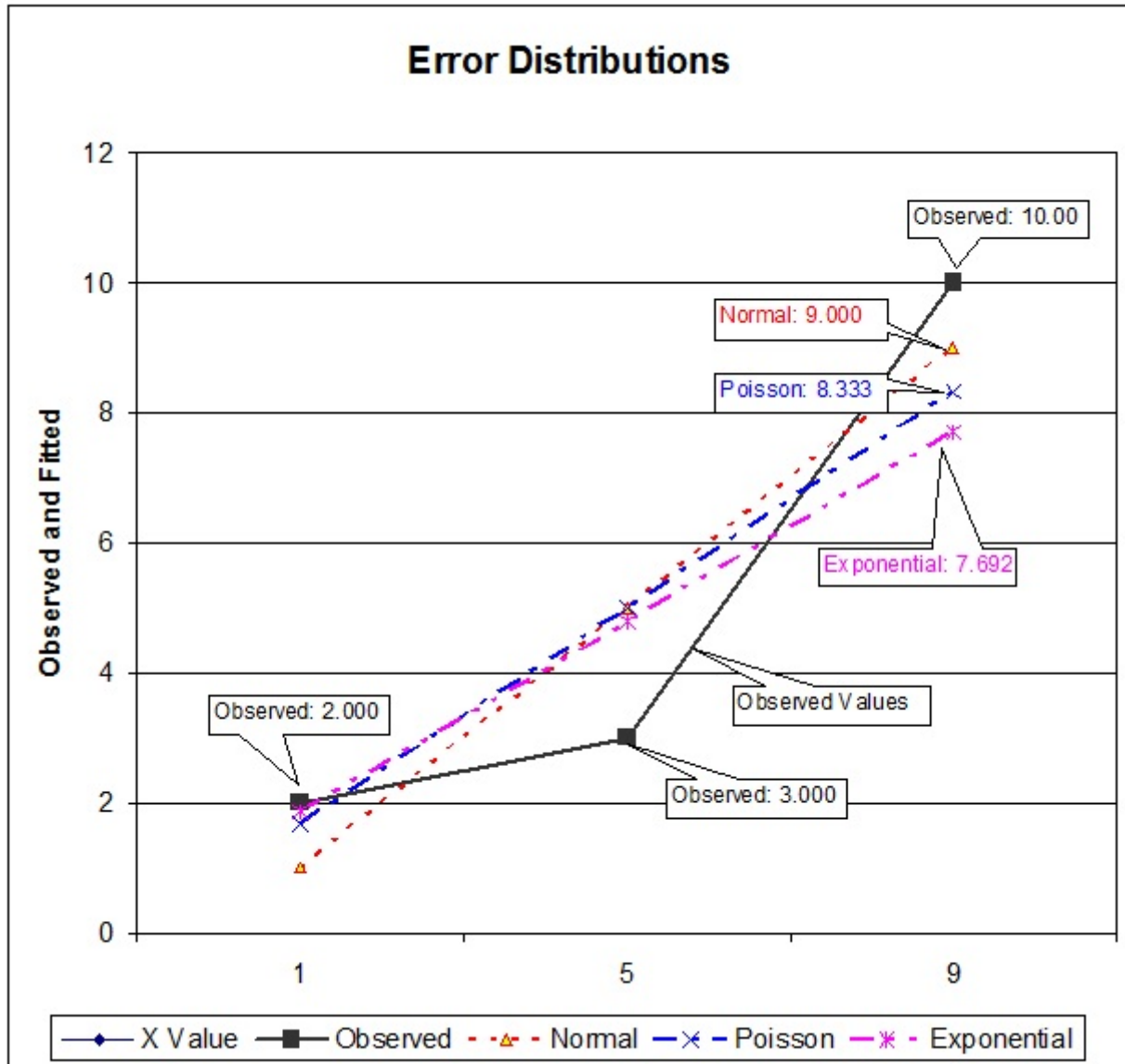
- Fitted Values: $7.692 / 1.866 = 4.122$.
- Variances are proportional to the squares of the fitted values: $4.122^2 = 16.991 \approx 17$
- Residuals: $2.308 / 0.134 = 17.224 \approx 17$.

Part J: We compute the residual divided by the variance of the error term, which is the square of the fitted value.

X	<i>Observed Value</i>	<i>Fitted Value</i>	<i>Residual</i>	<i>Residual Squared</i>	<i>Rsd / Var</i>
1	2	1.866	0.134	0.018	0.038
5	3	4.779	-1.779	3.163	-0.078
9	10	7.692	2.308	5.329	0.039
Total	15	14.336	0.664	8.510	0.000

The graphic below shows the three observed values and the fitted lines. The picture helps you understand the effect of the conditional distribution of the response variable.

- The three observed values do not fall on a straight line.
 - We assume the observed values are distorted by random fluctuation.
 - Each value is an expected value plus a random error.
 - The line of best fit depends on the variance of the error terms.



The relation of the variance to the fitted value rotates the fitted line.

- If the variances are equal at all points, the residuals are equal at $X=1$ and $X=9$ (Normal distribution with constant variance).
- If the variances are proportional to the fitted values, the residuals at $X=1$ and $X=9$ are proportional to the fitted values at those points (Poisson distribution).
- If the variances are proportional to the squares of the fitted values, the residuals at $X=1$ and $X=9$ are proportional to the squares of the fitted values (exponential distribution).

<i>X Value</i>	<i>Observed Value</i>	<i>Normal</i>		<i>Poisson</i>		<i>Exponential</i>	
		<i>Fitted</i>	<i>Residual</i>	<i>Fitted</i>	<i>Residual</i>	<i>Fitted</i>	<i>Residual</i>
1	2	1.000	1.000	1.668	0.332	1.866	0.134
5	3	5.000	-2.000	5.000	-2.000	4.779	-1.779
9	10	9.000	1.000	8.333	1.667	7.692	2.308
Total			0		0		0.663

You may want to run GLMs using R. The section below provides the code.

Excel's *SOLVER* add-in shows how GLMs derive the fitted values. Statistical software, such as SAS or R, use more efficient iterative techniques. R is perhaps the most powerful and flexible GLM package. The R code for this exercise is below.

The *glm* function runs a GLM of the response variable on the covariate. We use a Gamma distribution instead of an exponential distribution, which gives a slightly better fit, though the difference is less than 0.1%.

```
covariate <- c(1,5,9)
response <- c(2,3,10)
```

```
glm.pois.iden <- glm(response ~ covariate, family= poisson (link = "identity"))
glm.gamm.iden <- glm(response ~ covariate, family= Gamma (link = "identity"))
```

list element `[[3]]` gives the fitted values of the GLM.

```
glm.pois.iden[[3]]
glm.gamm.iden[[3]]
```

<i>R output</i>	<i>Observed Value</i>	<i>Normal</i>		<i>Poisson</i>		<i>Exponential</i>	
<i>X Value</i>		<i>Fitted</i>	<i>Residual</i>	<i>Fitted</i>	<i>Residual</i>	<i>Fitted</i>	<i>Residual</i>
1	2	1.000	1.000	1.667	0.333	1.865	0.135
5	3	5.000	-2.000	5.000	-2.000	4.779	-1.779
9	10	9.000	1.000	8.333	1.667	7.694	2.306
Total			0		0		0.6618

Exercise 1.5: Maximum likelihood estimation

We use maximum likelihood to fit a linear model to three points (1, 2), (5, 3), (9, 10).

We have a choice of three distributions for the error term:

- Normal distribution with a constant variance.
- Poisson distribution
- Exponential distribution

- A. Which distribution gives the lowest TSS (total sum of squares)?
B. Which distribution gives the lowest RSS (residual sum of squares)?

Part A: The mean Y value is $\frac{1}{3} \times (2 + 3 + 10) = 5$.

The total sum of squares (TSS) is $(2 - 5)^2 + (3 - 5)^2 + (10 - 5)^2 = 38$.

- The TSS measures the dispersion of the observed values from the overall mean, not from their fitted values.
- It does not depend on the distribution of the error terms about their means.

All three distributions have the same TSS.

Part B: We explain intuitively the relative sizes of the error sum of squares.

- For a normal distribution with a constant variance, the ordinary least squares estimators are the maximum likelihood estimators.
- The ordinary least squares estimators minimize the error sum of squares.

⇒ The RSS is lowest for the normal distribution with a constant variance.

Intuition: Suppose we have two points at opposite sides of the fitted line and we rotate the fitted center about its center.

- Moving one fitted value (FV_1) closer to its observed value (OV_2) moves the other fitted value (FV_2) farther from its observed value (OV_2).
- Linear regression (a normal distribution with a constant variance) does not favor any point above others.
- It gives the same weight to $X=1$ as to $X=9$.

If the distance from the mean to the observed value is d , the RSS is $N \times d^2$.

- A GLM moves one fitted value k units closer to its observed value at a cost of k units for another fitted value.
- The GLM changes its RSS from $2d^2$ to $(d - k)^2 + (d + k)^2 = 2d^2 + 2k^2$.