What type of regression model is this?

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

Standard Linear

What type of regression model is this?

$$Y_i \sim BER(\exp(\beta_0 + \beta_1 x_i) / [1+\exp(\beta_0 + \beta_1 x_i)])$$

Logistic

What type of distribution is this?

$$Y_i \sim POI(\exp(\beta_0 + \beta_1 x_i))$$

Poisson

What distribution does Y_i have in simple linear regression? Express as a function of x_i.

$$N\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$$

What distribution does Y_i have in logistic regression? Express as a function of x_i.

$$BER\left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}\right)$$

What distribution does Y_i have in Poisson regression? Express as a function of x_i.

$$POI\left(e^{\beta_0 + \beta_1 x_i}\right)$$

Suppose both predictor and outcome are interval variables. What type of regression might be appropriate?

Standard Linear

Suppose the predictor is nominal with 5 levels and the outcome is interval. What type of regression might be appropriate? How many binary 'dummy' predictors will be needed in the model?

Standard Linear

Suppose the outcome is binary but the predictor interval. What type of regression might be appropriate?

Logistic or Modified Poisson

Suppose the outcome and predictor are both binary. What type of regression might be appropriate?

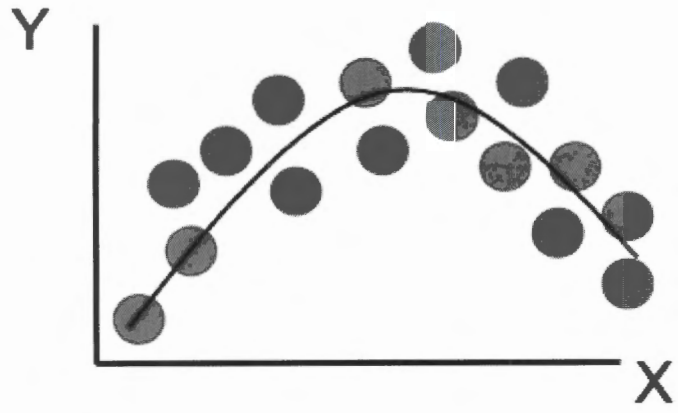
logistic or modified poisson

When using Poisson regression for binary outcomes, why must 'robust error estimates' be used?

The mean and variance of a Poisson distribution are the same.

This is not true with a binary (Bernoulli) outcome.

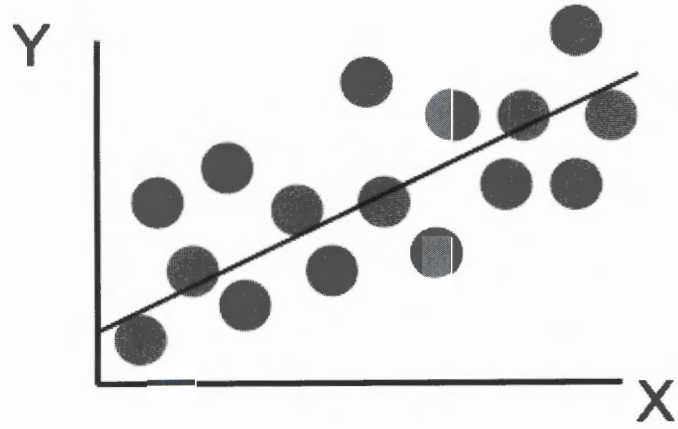
What regression model might be appropriate for this data? Specify the distribution of Y as a function of x.



$$Y_i = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon_i$$

$$Y_i \sim N(\beta_0 + \beta_1 x + \beta_2 x^2, \sigma^2)$$

What regression model might be appropriate for this data? Specify the distribution of Y as a function of x.



$$Y_i \sim N(\beta_0 + \beta_1 x, \sigma^2)$$

True or false: Predictors must be normally distributed for simple linear regression. False

True or false: A histogram of the outcomes should be approximately normally distributed in order to use simple linear regression?

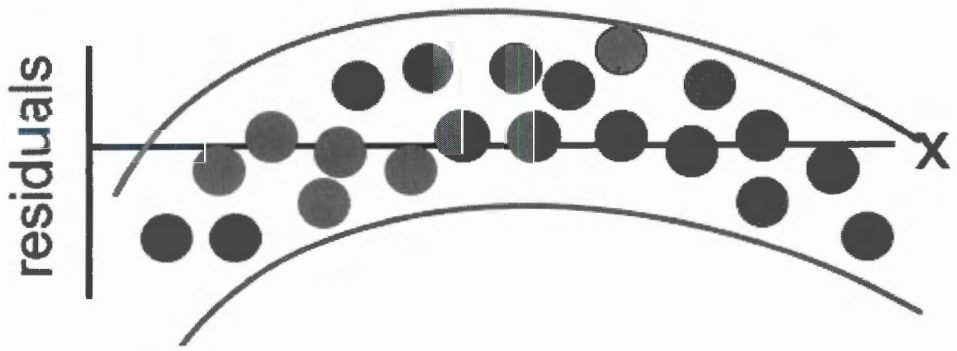
False

Name at least one way that the parameters could be estimated for a regression model.

Maximum Likelihood

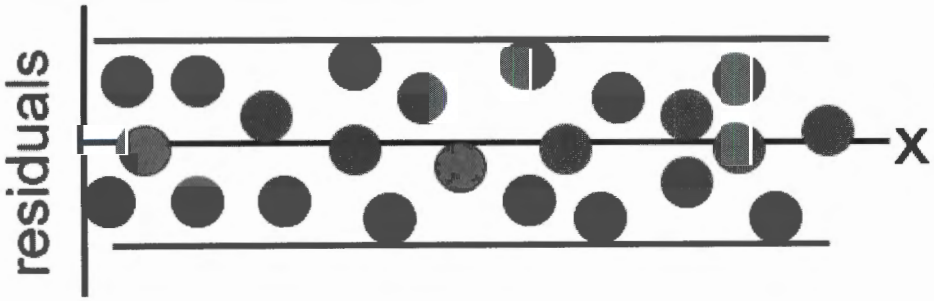
Why do researchers often not provide the estimate or confidence interval for the intercept in their models?

What is wrong with the following residuals for simple linear regression?



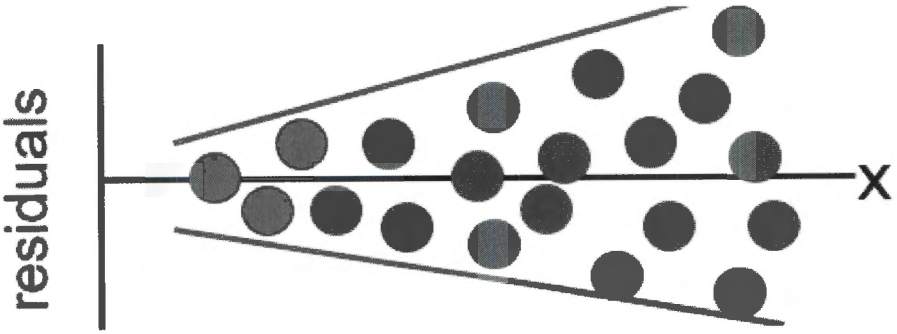
The relationship between x and y is not linear.

What is wrong with the following residuals for simple linear regression?



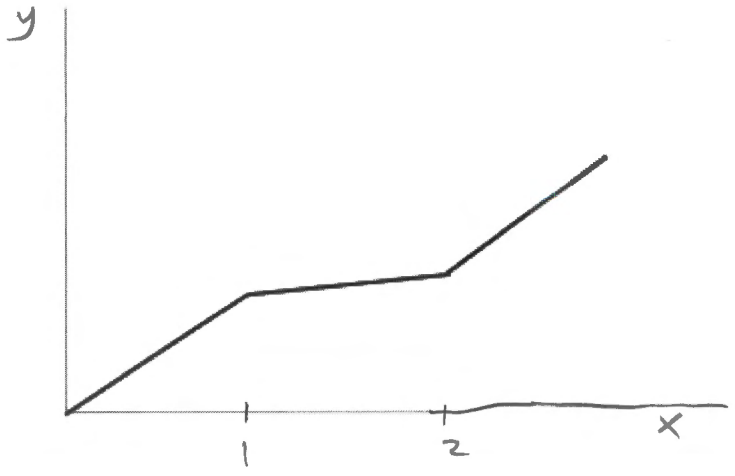
Nothing.

What is wrong with the following residuals for simple linear regression?



The variance is a function of x.

How might you model a relationship that looks like this?



$$Y_i = \beta_0 + \beta_2 x + \beta_3 (x-1) 1\{x > 1\} + \beta_4 (x-2) 1\{x > 2\} + \beta_4 (x-3) 1\{x > 3\} + \varepsilon_i$$

$$\varepsilon_i \sim iid \; N(0, \sigma^2)$$

What is meant by relative risk of an event? What is meant by odds ratio?

$$\text{Relative Risk} = \frac{p_1}{p_2}$$

$$\text{Odds Ratio} = \frac{\left(\frac{p_1}{1-p_1}\right)}{\left(\frac{p_2}{1-p_2}\right)}$$

When are relative risks and odds ratios similar? When are they quite different?

if p_1, p_2 are small, then odds ratio is similar to relative risk.

if p_1 or p_2 is large and the other is not close, they can be quite different.

The follow code was run:

```
proc genmod data = subjectlevel;
    model weight = age/ link = identity dist = normal;
run;
```

The following appears in the output:

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	3.1444	0.0470	3.0524	3.2365	4480.56	<.0001
AgeYears	1	-1.3909	2.0776	-5.4629	2.6810	0.45	0.5032
Scale	1	0.5703	0.0247	0.5239	0.6207		

What type of regression is this?

Standard Linear

What method was used to estimate the parameters?

MLE

Interpret the relationship of age to weight. Note that age is measured in years and weight in kg. Include a confidence interval in your interpretation.

The expected weight increases by -1.3909 $(-5.4629, 2.6810)$ for each 1 kg increase in weight.

The follow code was run:

```
proc genmod data = subjectlevel ;
    class Hypothermia(ref='No')/param = ref;
    model INDNewIntracranBleed (ref='No') = Hypothermia / link = logit dist = binomial;
run;
```

The following appears in the output: PROC GENMOD is modeling the probability that INDNewIntracranBleed='Yes'.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.4747	0.1634	-1.7950	-1.1543	81.40	<.0001
Hypothermia	Yes	1	1.0692	0.4848	0.1190	2.0194	4.86	0.0274
Scale		0	1.0000	0.0000	1.0000	1.0000		

What type of regression is this?

Logistic

What method was used to estimate the parameters?

MLE

Interpret the relationship of hypothermia to intracranial bleeding. Include a confidence interval in your interpretation.

The odds of bleeding are $e^{1.0692}$ $\left(e^{0.1190}, e^{2.0194}\right)$ times higher with hypothermia vs. without.

The follow code was run:

```
proc genmod data = subjectlevel;
    class Hypothermia(ref='No') StudySubjectID/param = ref;
    model INDNewIntracranBleed (ref='No')= Hypothermia / link = log dist = poisson;
    repeated subject = StudySubjectID;
run;
```

The following appears in the output:

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-1.6807	0.1330	-1.9414	-1.4201	-12.64	<.0001
Hypothermia	Yes	0.7645	0.3045	0.1677	1.3612	2.51	0.0120

What type of regression is this? Note that the repeated statement ensures that robust estimators are used for the standard error of the estimators.

Modified Poisson

Interpret the relationship of hypothermia to intracranial bleeding. Include a confidence interval in your interpretation.

The risk of bleeding is $e^{0.7645}$ $\left(e^{0.1677}, e^{1.3612}\right)$ times higher with hypothermia vs. without.

The follow code was run:

```
proc genmod data = subjectlevel;
    class Hypothermia(ref='No') StudySubjectID/param = ref;
    model INDNewIntracranBleed (ref='No')= Hypothermia LastBaselineFibrinogen / link = log dist = poisson;
    repeated subject = StudySubjectID;
run;
```

The following appears in the output:

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-0.7307	0.4982	-1.7072	0.2458	-1.47	0.1425
Hypothermia	Yes	0.6620	0.3270	0.0210	1.3029	2.02	0.0430
LastBaselineFibrinog		-0.0039	0.0026	-0.0090	0.0012	-1.51	0.1315

What type of regression is this? Note that the repeated statement ensures that robust estimators are used for the standard error of the estimators.

Modified Poisson

Interpret the relationship of hypothermia to intracranial bleeding. Include a confidence interval in your interpretation.

The risk of bleeding is $e^{0.6620}$ $(e^{0.0210}, e^{1.3029})$ times higher with hypothermia vs without, controling for fibrinogen;

Interpret the relationship of fibrinogen to intracranial bleeding. Note that fibrinogen is reported in mg/dL.

The risk of bleeding is $e^{-0.0039}$ $(e^{-0.0090}, e^{0.0012})$ times higher for each additional 1 mg/dL of fibrinogen. (controling for hypothermia use).

The follow code was run:

```
proc genmod data = subjectlevel;
    class Hypothermia(ref='No')/param = ref;
    model INDNewIntracranBleed (ref='No')= Hypothermia LastBaselineFibrinogen LastBaselineFibrinogen*Hypothermia / link = logit dist = binomial;
run;
```

The following appears in the output: PROC GENMOD is modeling the probability that INDNewIntracranBleed='Yes'.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.4666	0.5115	-1.4691	0.5359	0.83	0.3617
Hypothermia	Yes	1	4.0024	2.1521	-0.2156	8.2204	3.46	0.0629
LastBaselineFibrinog		1	-0.0039	0.0023	-0.0083	0.0005	2.97	0.0847
LastBasel*Hypothermi	Yes	1	-0.0140	0.0096	-0.0327	0.0048	2.13	0.1440
Scale		0	1.0000	0.0000	1.0000	1.0000		

What type of regression is this?

Logistic

Interpret the relationship of fibrinogen to intracranial bleeding. Note that fibrinogen is reported in mg/dL. ~~Include a confidence interval in your interpretation.~~

If hypothermia is used, then the odds of bleeding is $e^{-0.0039-0.0140}$ times higher for each additional 1 mg/dL of fibrinogen.

If hypothermia is <u>not</u> used, then the odds of bleeding is $e^{-0.0039}$ times larger for each additional 1 mg/dL of Fibrinogen.

The following code was run:

```
proc genmod data = subjectlevel;
    class Hypothermia(ref='No') indicationECMO StudySubjectID/param = ref;
    model INDNewIntracranBleed (ref='No')= IndicationECMO / link = log dist = poisson;
    repeated subject = StudySubjectID;
run;
```

The following output was generated from the above code:

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-1.5208	0.1539	-1.8224	-1.2192	-9.88	<.0001
IndicationECMO	Cardiac	-0.1678	0.2676	-0.6924	0.3568	-0.63	0.5307
IndicationECMO	ECPR	-0.2710	0.4817	-1.2151	0.6731	-0.56	0.5737

What type of regression is this?

Modified Poisson

Interpret the relationship between the indication for ECMO and intracranial bleeding. Note that indication for ECMO is a nominal variable with three levels: Respiratory, Cardiac, and eCPR.

The risk of intracranial bleeding is $e^{-0.1678}$ times higher with Cardiac indication vs. respiratory indication.

The risk of intracranial bleeding is $e^{-0.2710}$ times higher with eCPR indication vs. respiratory indication.

~~What is the distribution of intracranial hemorrhage (confirmation of indication for ECMO)?~~