Lecture 4. Logistic Regression

Suzer-Gurtekin

January 2025

Action Plan (I)

- Logistic regression
    - Rationale behind it
    - Interpretation of coefficients
    - Inference
- While you can use R functions to fit and test your models, you should include:
    - Explicit model definition (including variable names, and categories) – see in-class exercise 3
    - R code
    - Output interpretation
- Q&A

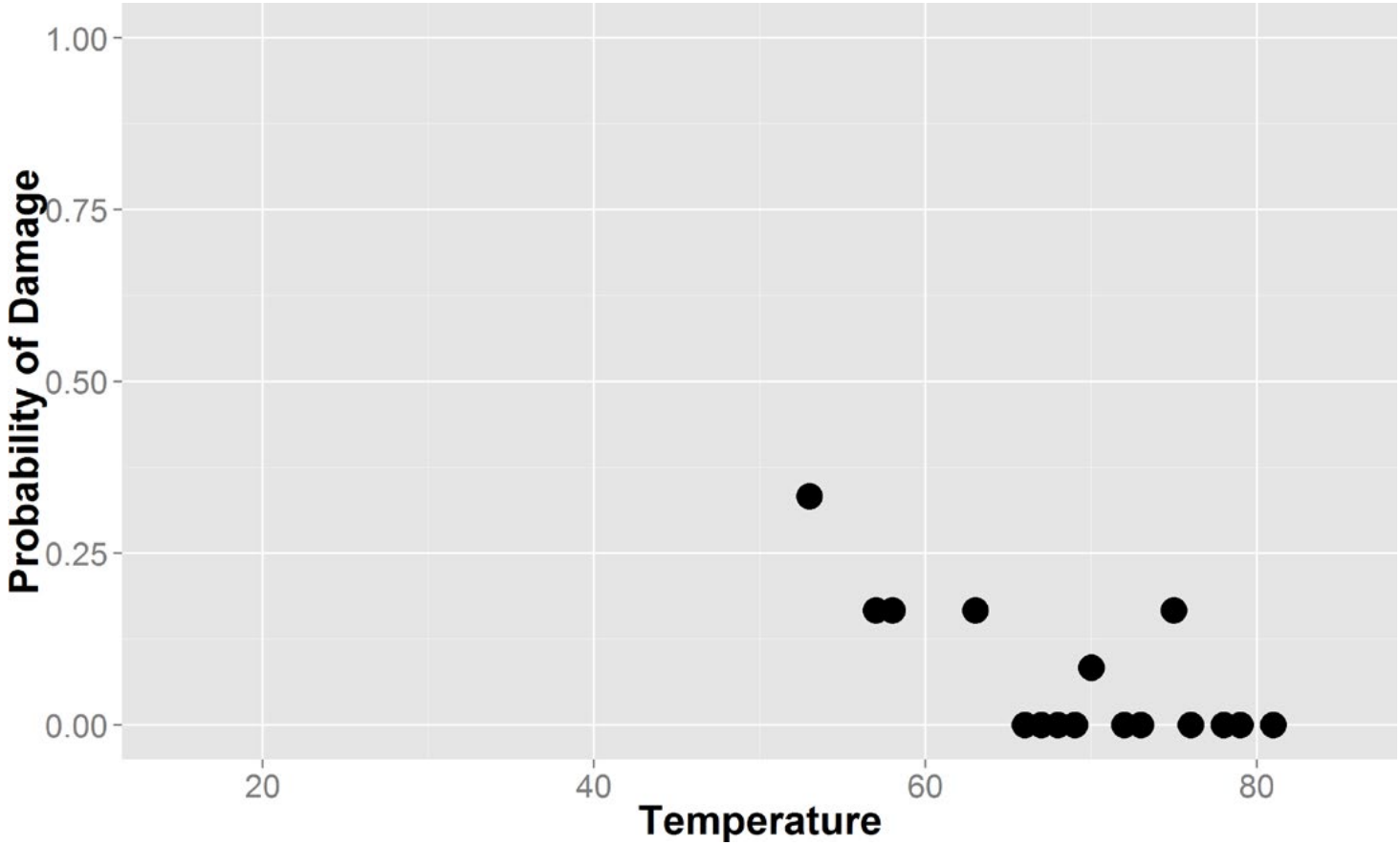
Overview

Example 1: O-rings on the Space Shuttle

- O-rings fail at higher rates in colder temperatures
- There are 6 O-rings on each shuttle
- There were 23 missions prior to Challenger
- Could the Challenger explosion have been predicted?

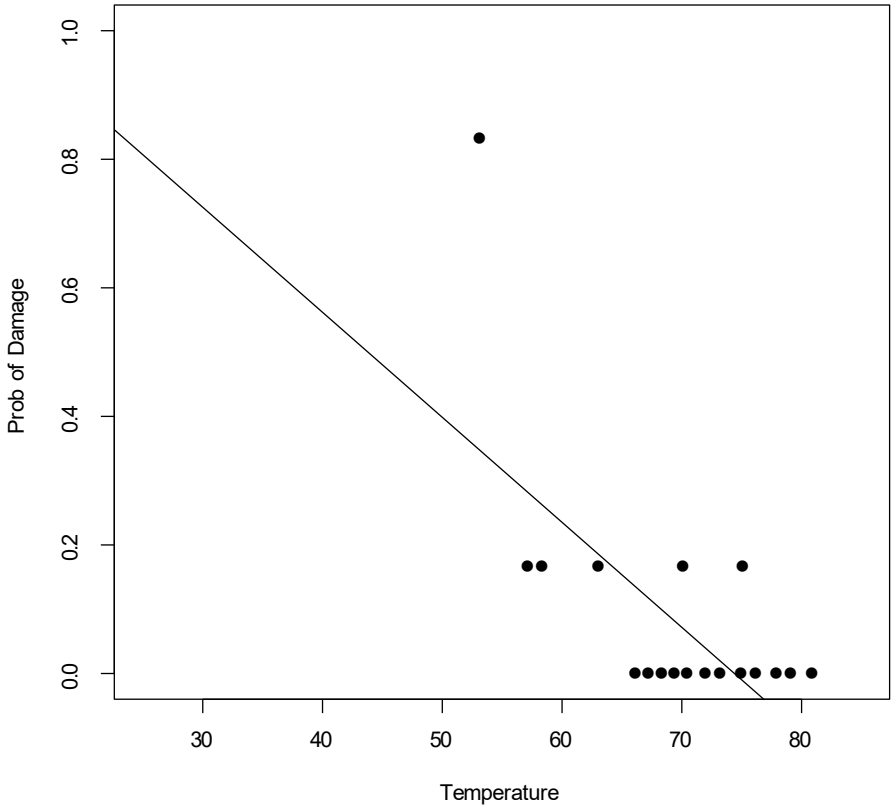
Example 1: O-rings on the Space Shuttle

First, plot the failure rates by temperature:



Example 1: O-rings on the Space Shuttle

It looks like there might be a relationship. Let's regress failure rate on temperature. Here's the resulting regression line:



Linear Model and Binary Outcome

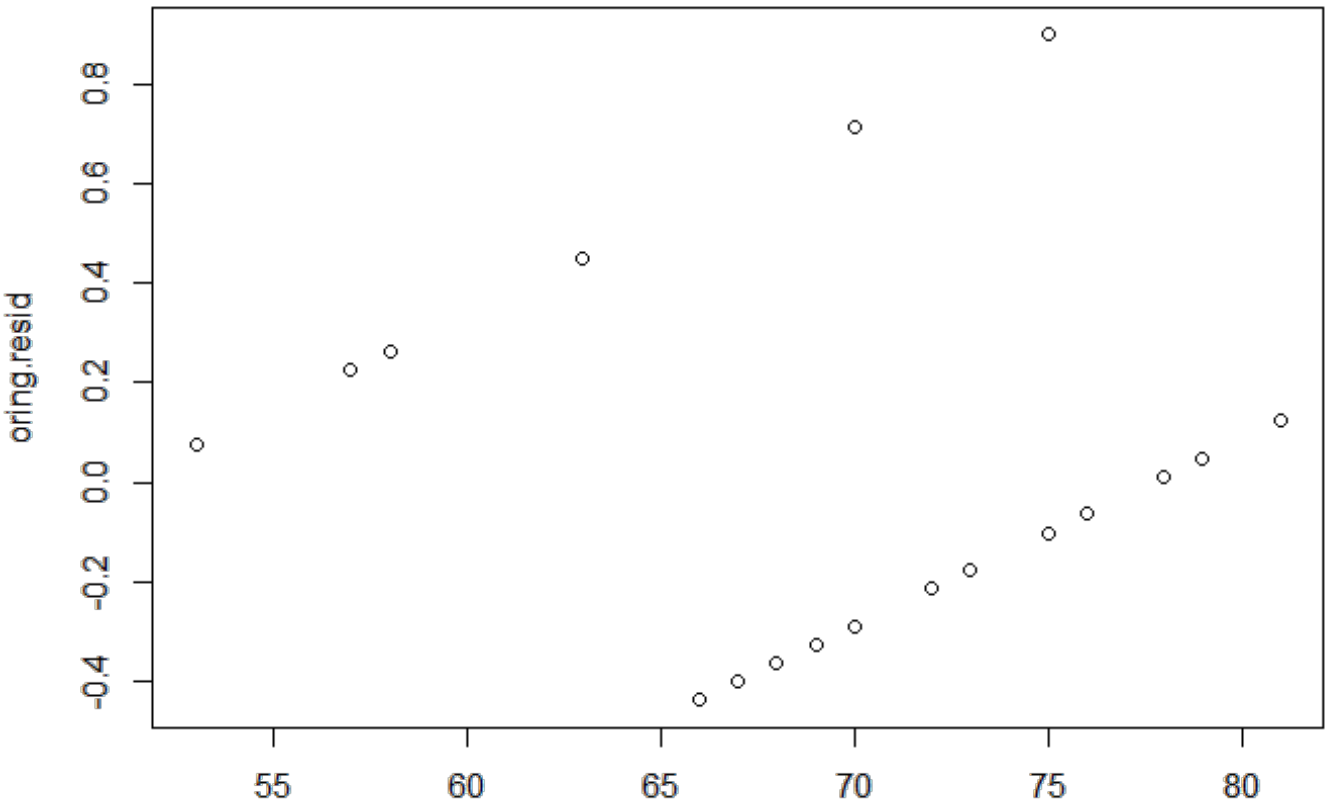
In some cases, the linear model might not produce a bad fit, but there are some issues.

1. Linear regression assumes residuals are normally distributed.
2. Linear regression also assumes constant variance.
3. Linear model can produce predicted values outside the restricted range of a probability (0,1).

Violations of assumptions mean linear model won't always/usually be appropriate.

Linear Model and Binary Outcome

Assume that residuals are normally distributed. These are the residuals from a linear model $DAMAGE = \beta_0 + \beta_1 \times TEMP$:



Linear Model and Binary Outcome

Recall the binomial distribution:

$$P(Y_i = y_i) = \frac{n_i!}{y_i!(n_i - y_i)!} p_i^{y_i} (1 - p_i)^{n_i - y_i} \quad (1)$$

This distribution has mean $n_i p_i$ and variance $n_i p_i (1 - p_i)$. The mean and the variance are linked in this case – nonconstant. Violation of second assumption.

Linear Model and Binary Outcome

Let's say we also have q predictors $(x_{i1}, \ldots, x_{iq})$. We want to use these to predict p_i.

For the reasons described earlier, that may not work.

Question:

Is it possible to define a function, $g(p_i)$ for which the linear regression model will work?

Logistic Regression

This is the basic idea behind logistic regression (and many methods under the heading of generalized linear models [GLMs]).

We need some function $\eta_i = g(p_i)$ such that η_i can be predicted in a linear model.

It must also satisfy $0 \leq g^{-1}(\eta) \leq 1$.

The most common function is called the logit function: $\eta = log\ (\frac{P}{1-P})$

Logistic Regression

Table: Relationship of Odds and Log-odds to Probabilities

Probability	Odds	Log-Odds
0	0	$-\infty$
0.1	0.111	-2.197
0.2	0.25	-1.386
0.3	0.429	-0.847
0.4	0.667	-0.405
0.5	1	0
0.6	1.5	0.405
0.7	2.333	0.847
0.8	4	1.386
0.9	9	2.197
1	∞	∞

Logistic Regression

Logistic regression fits the following model:

$$\eta = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_q x_{iq}$$

$$log \quad \frac{p}{1-p} = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_q x_{iq}$$

$$Pr(Y_i = 1) = logit^{-1} \left(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_q x_{iq} \right)$$

$$Pr(Y_i = 1) = \frac{1}{1 + e^{-\left(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_q x_{iq} \right)}}$$

Logistic Regression

The R code is straightforward. We'll start with the basics and add to this as we go:

```
logitmod<-glm(damage ~temp, family=binomial,
data=spacshu)
```

Logistic Regression

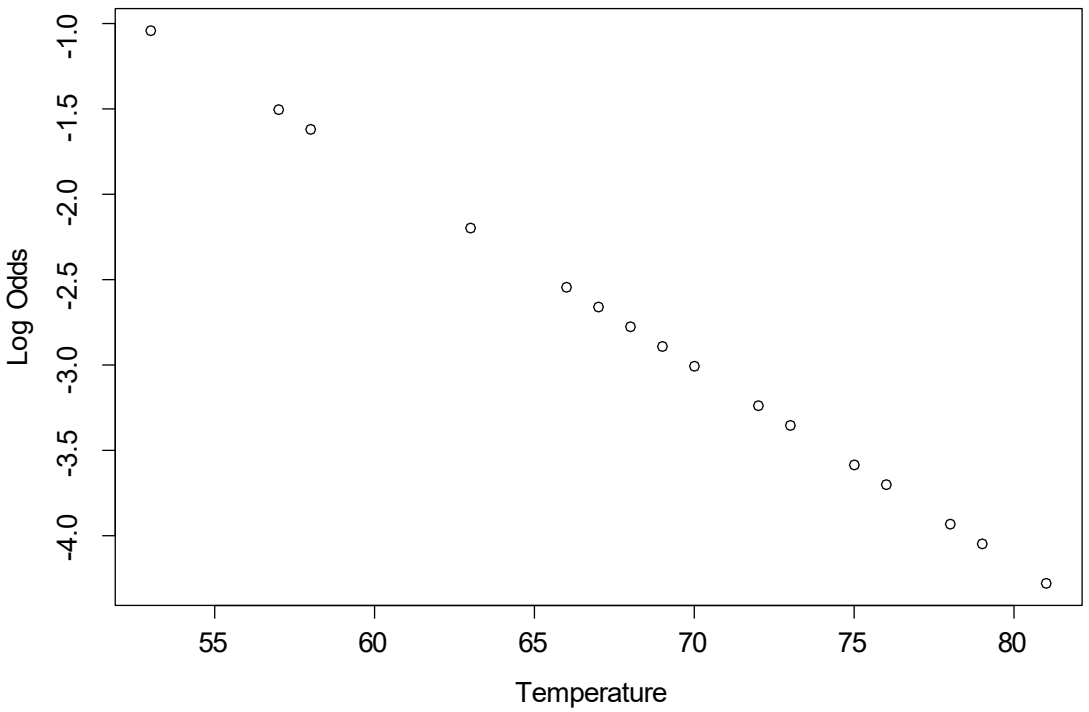
If we fit the model $\eta = \beta_0 + \beta_1 x_{i1}$, where $\eta = log\ (\frac{p}{1-p})$ and x is the temperature, we get the following:

```
Coefficients:
(Intercept)          temp
5.0850               -0.1156
```

In this case $\eta = 5.0850 - 0.1156 * TEMP$.

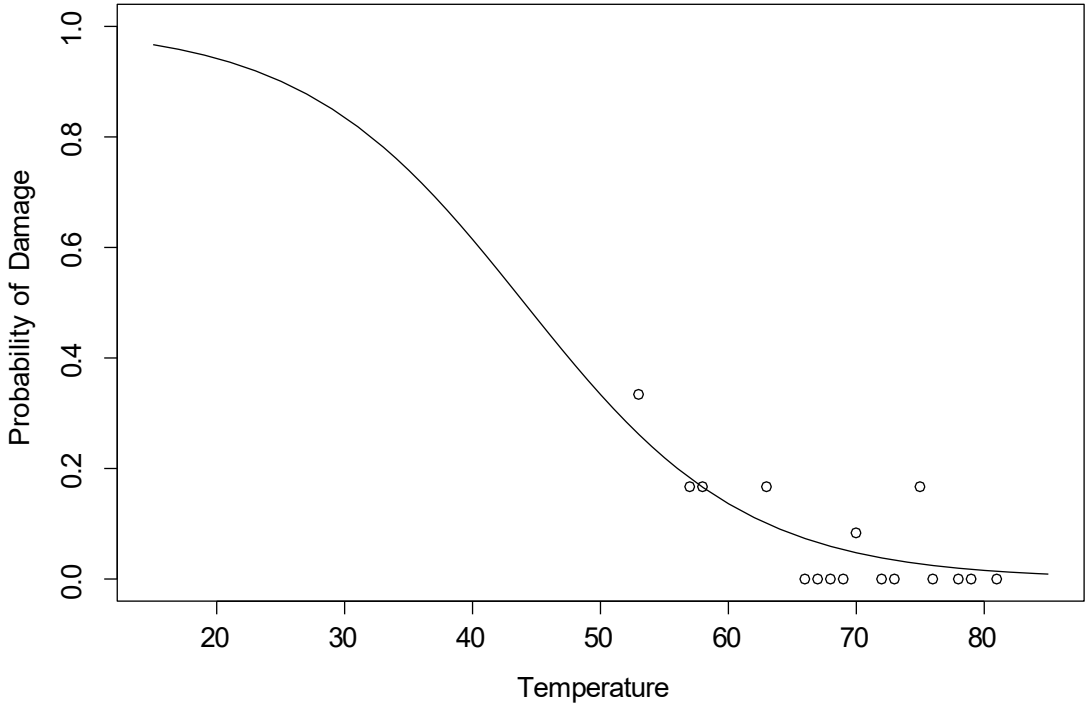
Logistic Regression

Let's plot the predicted values from this model.



Logistic Regression

We could also convert this to the probability scale:



Logistic Regression

It is possible to use many of the techniques that you learned last semester to predict the logit in this setup.

- Transformations
- Dummy variables
- Interactions

Interpretation of Coefficients

But, we are working on the scale of the logit. This means that the results need careful interpretation.

Recall the $\frac{p}{1-p}$ is called the odds. So η is actually a log of the odds.

We can write our model:

$$log\,(odds) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_q x_{iq}$$

Odds Ratio

Within row 1 the *odds* (***not* odds ratio**) that the response is in column 1 instead of column 2 is defined as:

$$Odds_1 = \frac{\pi_{1|1}}{\pi_{2|1}} = \frac{\pi_{11}}{\pi_{12}}$$

$$\frac{probability\ of\ you\ are\ in\ column\ 1\ given\ that\ you\ are\ in\ row\ 1}{probability\ of\ you\ are\ in\ column\ 2\ given\ that\ you\ are\ in\ row\ 1}$$

From our example, this could be written:

$$\frac{\Pr(D|M)}{\Pr(\overline{D}|M)} = \frac{\Pr(D|M)}{1-\Pr(D|M)}$$

	Yes	No	Row Total
Male	π_{11}	π_{12}	π_{1+}
Female	π_{21}	π_{22}	π_{2+}
Column Total	π_{+1}	π_{+2}	

Continuing the example, the odds that a man will have the disease are $\frac{.3}{1-.3} = .43$. For women, this odds are $\frac{.2}{1-.2} = .25$

Interpretation of Coefficients

$$log\,(odds) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_q x_{iq}$$

$$odds = exp\,(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_q x_{iq})$$

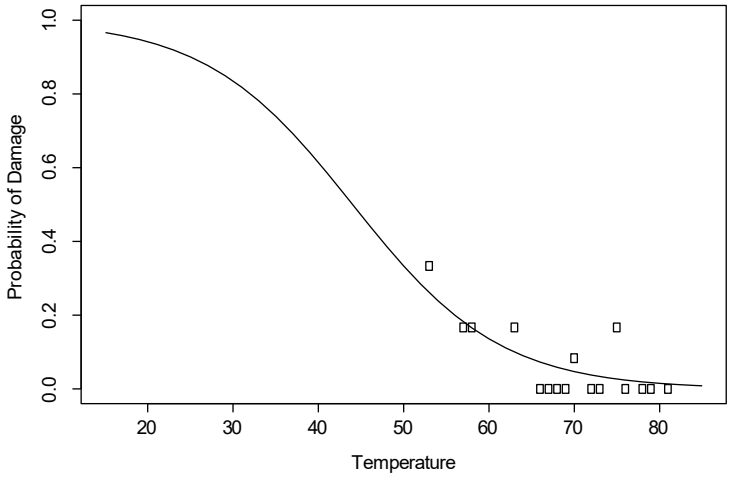
In this case, β_1 can be interpreted as meaning a 1-unit increase in x_{i1} with all other x's being held constant increases the log-odds of success by β_1.

Therefore, e^{β_1} can be interpreted as the increase in the odds-ratio that would result from a 1-unit increase in x_{i1} with all other x's being held constant.

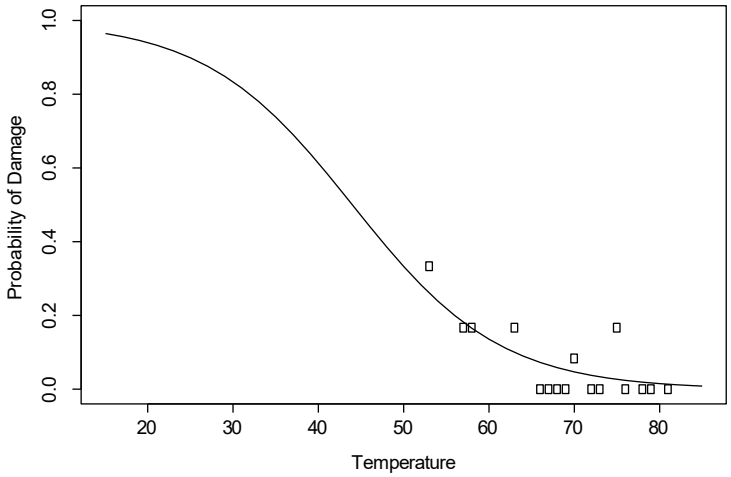
Interpretation of Coefficients

One issue is that people want to interpret results on the probability scale. This won't work very easily with logistic regression.

Recall that the inverse-logistic function is curved:



Interpretation of Coefficients



The steepest change occurs at the middle of the curve.

Interpretation of Coefficients

Example Points

logit(0.50)=0 and logit(0.60)=0.4. At this point, adding 0.4 on the logit scale increases the probability 10%.

logit(0.90)=2.2 and logit(0.93)=2.6. At this point, adding 0.4 on the logit scale increases the probability 3%.

Interpretation of Coefficients

One method is to evaluate at the mean of the predictor variable of interest.

In the O-ring example, the mean temperature is 69.57. Therefore, the probability of O-ring failure at that temperature is

$$\frac{1}{1+e^{-(5.0850-0.1156*69.57)}} = 0.05.$$

Interpretation of Coefficients

Another method is to evaluate at several points. This assumes that all the values of all other predictors are held constant (if we had more x's).

Table: Pr(Failure) at Different Temperatures

Temperature	Logit	Pr(Failure)	Increase
69.57	-2.96	0.05	
59.57	-1.80	0.17	0.11
49.57	-0.65	0.52	0.36

Interpretation of Coefficients

The 'divide by 4 rule'.

The slope of the curve is steepest at its center point. At this point, $\beta_0 + \beta_1 x_{i1} = 0$ so that $logit^{-1}(0) = 0.5$. The slope of the curve at that point is $\beta_1/4$.

Therefore, the maximum difference in probability corresponding to a 1-unit increase in x is $\beta_1/4$. This maximum occurs where the probabilities are close to 0.5. In our example $-0.1156/4 = -0.0289$.

Interpretation of Coefficients

Why does the 'divide by 4 rule' work?

The slope of the curve reaches its maximum at $\beta_0 + \beta_1 * x_{i1} = 0$. At this point, $logit^{-1}(\beta_0 + \beta_1 * x_{i1}) = 0.5$.

You can use the first derivative of the logit function with respect to β_1 to get an estimate of the maximum slope:

$$\beta \frac{e^0}{(1 + e^0)^2} = \beta \frac{1}{(1 + 1)^2} = \frac{\beta}{4} \quad (2)$$

Interpretation of Coefficients

Another already mentioned interpretation of the coefficients in a logistic regression model is as odds ratios.

Recall that the log of the odds ratio has a symmetric distribution.

```
> exp(cbind(coef(logitmod), confint(logitmod)))
Waiting for profiling to be done...
                            2.5 %      97.5 %
(Intercept)  161.5762606 0.3641231 7.207593e+04
temp           0.8908304 0.8087817 9.758269e-01
```

Interactions

It is possible to include interactions in logistic regression models.

These interactions need careful interpretation.

Interactions

Example from Chen (2003).

Variable (Reference)	OR
Age, y(<30)	2.5
Married	1.1
No. of children (>=3)	0.8
Contraceptive use	0.1
Aware of vertical transmission	1.7
Planned pregnancy	1.7
HIV infected	3.0
No. of HIV-related symptoms (>=2)	0.9
Intention x infected	0.1
Frequency of sexual activity	0.9
Treatment group	1.6

Interactions

Calculating odds ratios in the presence of interactions. For women who were not planning a pregnancy, the OR for those with HIV infection vs. those without:

$$\hat{OR}_1 = \frac{\text{Odds for } (I_{planned} = 0 \text{ and } I_{infected} = 1)}{\text{Odds for } (I_{planned} = 0 \text{ and } I_{infected} = 0)}$$
$$= exp(\hat{\beta}_{infected})$$
$$= exp(1.0986) = 3.0$$

Interactions

For women with a planned pregnancy, the OR for those with HIV infection vs. those without:

$$\hat{OR}_1 = \frac{\text{Odds for } (I_{planned} = 1 \text{ and } I_{infected} = 1)}{\text{Odds for } (I_{planned} = 1 \text{ and } I_{infected} = 0)}$$
$$= exp(\hat{\beta}_{infected} + \hat{\beta}_{plannedxinfected})$$
$$= exp(1.0986 - 2.3026) = 0.3$$

Summary: Write out the OR of interest in this including both of the interacted variables and derive the OR appropriately.

In class Exercise

TABLE 2—Univariate Predictors of Pregnancy Status at Follow-up

Variable (Reference)	Women (n = 808)			Men (n = 826)		
	Pregnant, % (n)	Not Pregnant, % (n)	OR (95% CI)	Pregnant, % (n)	Not Pregnant, % (n)	OR (95% CI)
Age, y (< 30)	88 (52)	72 (418)	2.9 (1.3, 6.5)*	68 (42)	60 (405)	1.4 (0.8, 2.5)
Married/cohabiting	67 (39)	58 (329)	1.5 (0.9, 2.7)	66 (40)	46 (307)	2.3 (1.3, 3.9)*
No. of children (≥ 3)	12 (7)	26 (151)	0.4 (0.2, 0.9)*	12 (7)	18 (120)	0.6 (0.3, 1.4)
Religion (Christian)	71 (42)	72 (419)	1.0 (0.5, 1.7)	68 (42)	59 (399)	1.5 (0.8, 2.5)
Contraceptive use	17 (10)	56 (326)	0.2 (0.1, 0.3)*	31 (19)	56 (381)	0.3 (0.2, 0.6)*
Planned pregnancy	41 (24)	27 (155)	1.9 (1.1, 3.2)*	35 (21)	18 (116)	2.5 (1.4, 4.5)*
Aware of vertical transmission	85 (50)	80 (454)	1.4 (0.7, 2.9)	87 (54)	68 (450)	3.2 (1.5, 6.8)*
HIV infected	27 (16)	25 (142)	1.1 (0.6, 2.1)	13 (8)	11 (75)	1.2 (0.5, 2.6)
No. of HIV symptoms (≥ 2)	36 (21)	35 (201)	1.0 (0.6, 1.8)	34 (21)	31 (208)	1.2 (0.7, 2.0)
Recruitment site (Kenya)	41 (24)	53 (309)	0.8 (0.5, 1.4)	47 (29)	50 (338)	0.9 (0.5, 1.5)

Note. Reference = reference group for odds ratios; OR = odds ratio; CI = confidence interval.
*$P < .05$.

Reconstruct a two by two table for HIV infected by Being Pregnant for Men and Women independently using Table 2 from Forstyh et al. 2002 and interpret ORs reported in Table 2

In class Exercise

TABLE 3—Multiple Logistic Regression of Predictors of Time 2 Pregnancy for Women (n = 639)

Variable (Reference)	OR (95% CI)
Age, y (< 30)	2.5 (1.0, 6.5)*
Married	1.1 (0.5, 2.1)
No. of children (≥ 3)	0.8 (0.3, 2.1)
Contraceptive use	0.1 (0.1, 0.3)**
Aware of vertical transmission	1.7 (0.8, 3.8)
Planned pregnancy	1.7 (0.8, 3.5)
HIV infected	3.0 (1.3, 7.0)**
No. of HIV-related symptoms (≥ 2)	0.9 (0.4, 1.6)
Intention × infected	0.1 (0.0, 0.4)**
Frequency of sexual activity	0.9 (0.4, 1.6)
Treatment group	1.6 (0.8, 3.2)

Note. Reference = reference group for comparisons; OR = odds ratio; CI = confidence interval.
*$P < .10$; **$P < .05$.

TABLE 5—Multiple Logistic Regression of Predictors of Partner Pregnancy for Men (n = 586)

Variable (Reference)	OR (95% CI)
Age, y (< 30)	2.1 (1.0, 4.3)*
Married	2.0 (1.0, 4.1)*
No. of children (≥ 3)	0.7 (0.2, 1.7)
Contraceptive use	0.2 (0.1-0.5)*
Aware of vertical transmission	4.2 (1.7, 10.1)*
Planned pregnancy	1.2 (0.6, 2.4)
HIV infected	1.6 (0.5, 5.1)
No. of HIV-related symptoms (≥ 2)	1.1 (0.6, 2.1)
Intention × infected	0.5 (0.1, 3.4)
Frequency of sexual activity	0.7 (0.3, 1.3)
Treatment group	1.9 (1.0, 3.5)*

Note. Reference = reference group for comparisons; OR = odds ratio; CI = confidence interval.
*$P < .05$.

Recompute the odds ratio for women who were planning a pregnancy versus those who were not is for women who were not HIV infected using Tables 3 and 5

Group Assignments – Please join the breakout rooms as follows:

Group	Student
1	Einolf, Zach Scott
1	Fan, Zhaoyun
1	Mishra, Rohin Prem
1	DesJardins, Grace
2	Adeniyi, Kehinde
2	Lugu, Nicholas Reign
2	LU, Aria
2	Gunderson, Jeremy
3	Wenner, Theodore D
3	Zhou, Zhenjing
3	Kim, Jay
3	Bei, Rongqi
4	Beshaw, Yael Dejene
4	Hoglund, Quentin Michael
4	Jiang, Yujing
4	Jiang, Weishan
5	Popky, Dana
5	Sani, Jamila
5	O'Connell, Greg Al
5	Saucedo, Valeria Castaneda

Group	Student
6	Hussein, Aya Moham
6	Zou, Jianing
6	Wang, Zixin
6	Chakravarty, Sagnik
7	Valmidiano, Megan
7	Glidden, Sarah Acton
7	Sun, Yao
7	Blakney, Aaron
8	Xu, Kailin
8	Linares, Kevin
8	Odei, Doris
8	Nana Mba, Line
9	Zhou, Huan
9	Meng, Lingchen
9	Lin, Xinyu
9	Ge, Feiran
10	Liu, Xiaoqing
10	Lu, Angelina
10	Baez-Santiago, Felix
10	Ma, Ruisi

Group	Student
11	Ding, Yuchen
11	Shrivastava, Namit
11	Kakiziba, Johnia Johansen
11	Cranmer, Evan Koba

Reference vs Effect/Sum Coding

There are alternative **parameterizations** for categorical predictors when used in logistic regression (and elsewhere).

These parameteriziations don't change predicted values, but the estimated coefficients are different and, therefore, require different interpretations.

Two different paramaterizations we will look at:

1. Reference coding
2. Effect or sum coding

Reference vs Effect/Sum Coding

The different codings are based on different **contrasts**.

Reference coding treats one category as a reference group.

This group's probability is estimated via the intercept.

Other groups have coefficients representing differences from the reference group.

Reference vs Effect/Sum Coding

Let's consider an example. The probability of a low birth weight baby.

Using `RACE` with three categories as a predictor.

The reference coding for this:

```
> contrasts(lowbwt$RACE)
  2 3
1 0 0
2 1 0
3 0 1
```

Reference vs Effect/Sum Coding

Sum or **Effect Coding** uses a contrast that compares each group to a "grand mean".

The constraint is that the coefficients must sum to 0.

Reference vs Effect/Sum Coding

Back to the same example.

Using `RACE` with three categories as a predictor.

The sum or effect coding for this:

```
> contrasts(lowbwt$RACE)
     [,1] [,2]
1     1    0
2     0    1
3    -1   -1
```

Reference vs Effect/Sum Coding

The estimates from the example:

Reference Coding						Effect/Sum Coding					
RACE	RACE2	RACE3				RACE	RACE1	RACE2			
1	0	0				1	1	0			
2	1	0				2	0	1			
3	0	1				3	-1	-1			
	Estimate	Std. Error	z value	Pr(>|z|)			Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.155	0.2391	-4.83	1.36E-06	***	(Intercept)	-0.6613	0.1759	-3.759	0.000171	
RACE1	0					RACE1	-0.4937	0.2236	-2.207	0.02728	
RACE2	0.8448	0.4634	1.823	0.0683	.	RACE2	0.3511	0.2889	1.215	0.224223	
RACE3	0.6362	0.3478	1.829	0.0674		RACE3	0.1426				
Logit						Logit					
RACE=1	-1.155		-1.155	0.239577		RACE=1	-0.6613+(-0.4937)		-1.155	0.239577	
RACE=2	-1.155+0.8448		-0.3102	0.423066		RACE=2	-0.6613 -0.3511		-0.3102	0.423066	
RACE=3	-1.155+0.6362		-0.5188	0.373133		RACE=3	-0.6613 - (-0.4937) - 0.3511		-0.5187	0.373156	
						OR RACE=3	-0.6613+0.1426		-0.5187		

Model Fit

For two-way tables we considered the deviance, G^2, which was a measure of how far the observed data departed from the expected data.

The saturated model has 0 for the deviance.

Model Fit

Recall that G^2 is based on the ratio of the likelihoods for the observed and expected data.

To apply that idea here, we need to start from the likelihood.

Model Fit

We can use this likelihood to test model fit.

First define $G_p = -2logL(\hat{\beta})$ as the likelihood for the model with p coefficients.

Then define $G_k = -2logL(\hat{\beta})$ as a reduced model with k coefficients.

Then the likelihood ratio test can be written as $-2log\frac{L(\hat{\beta}_k)}{L(\hat{\beta}_p)}$.

This ratio has a x^2 distribution with $p - k$ degrees of freedom.

Model Fit

This can also be written as $G_k - G_p \sim x^2_{p-k}$.

This should allow us test a model vs an alternative.

The conclusion will be whether the full model contains additional information beyond the reduced model.

Model Fit

We can use this method to compare our hypothesized model to a null model.

The null model has only the intercept (the mean).

Do our covariates add anything to our knowledge of the probabilities?

We could also ask, can we successfully differentiate cases in our sample into those with different probabilities?

Model Fit

Let's go back to our O-ring example.

```
> anova(logitmod,test='LRT')
Analysis of Deviance Table

Model: binomial, link: logit

Response: damage

Terms added sequentially (first to last)


     Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                   137     66.540
temp  1    6.144       136     60.396  0.01319 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It appears that temperature adds to our understanding of the failure of O-rings.

Model Fit

Similar output directly from `glm` :

```
> logitmod

Call:  glm(formula = damage ~ temp, family = binomial,

Coefficients:
(Intercept)         temp
     5.0850      -0.1156

Degrees of Freedom: 137 Total (i.e. Null);  136 Residu
Null Deviance:      66.54
Residual Deviance: 60.4   AIC: 64.4
```

Model Fit

This gives us a means to assess model fit and compare alternative models. We may also want to assess each coefficient in the model.

Fortunately, we can use some approximations here.

Model Fit

```
> Anova(logitmod, type="II", test="Wald")
Analysis of Deviance Table (Type II tests)

Response: damage
     Df  Chisq    Pr(>Chisq)
temp  1 6.0435    0.01396 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0
```

Model Fit

There is an assumed linear relationship between continuous predictors and the logit.

This assumption can be checked by reviewing *empirical logit plots*.

1. Calculate probability in binned values of the predictor.
2. Take the logit of these probabilities
3. Plot them. Is the relationship linear?

Model Fit

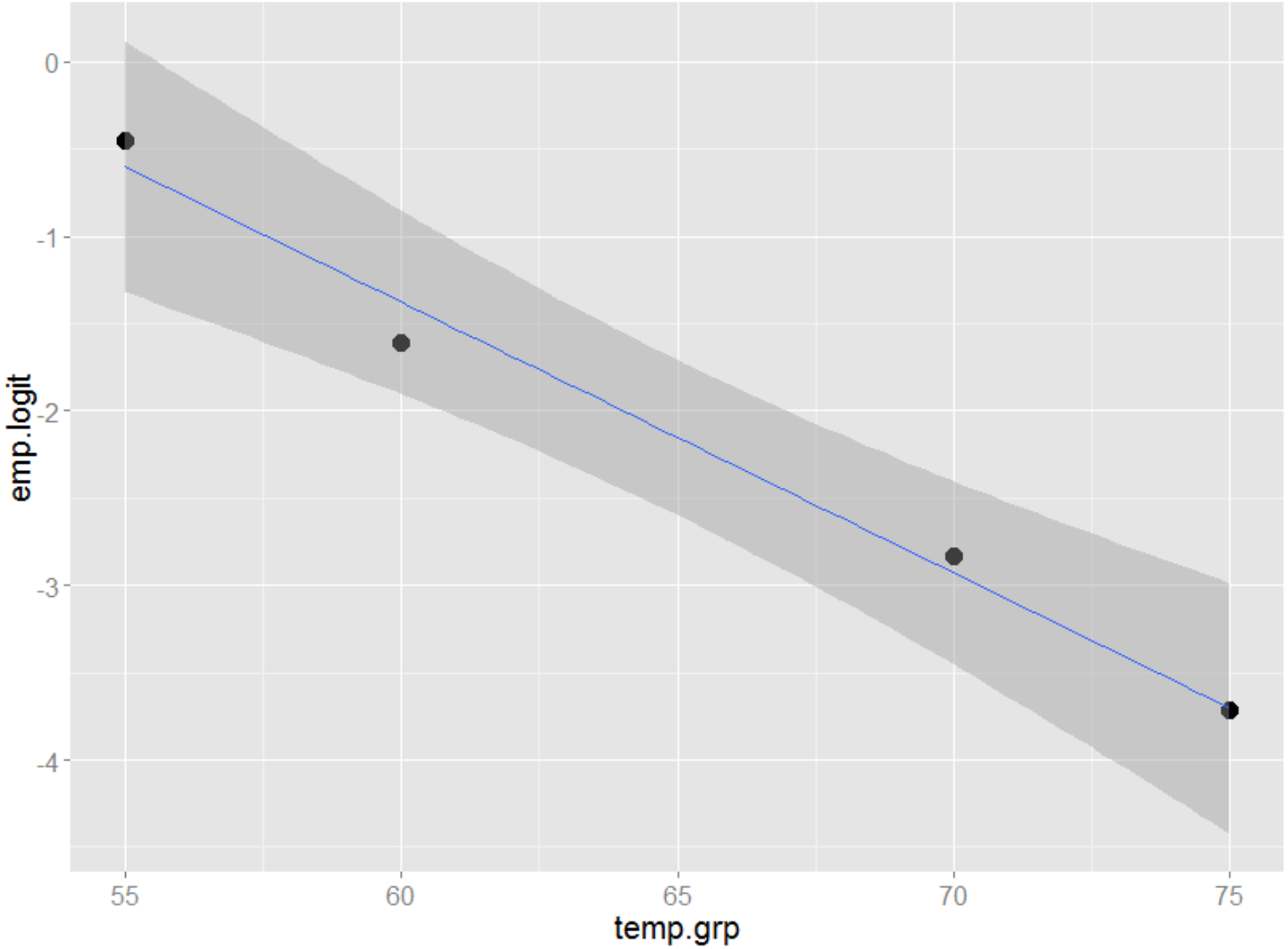
```
orings2<-orings
orings2$temp.grp[orings2$temp< 60] <- 55
orings2$temp.grp[orings2$temp >= 60 & orings2$temp <65 ] <- 60
orings2$temp.grp[orings2$temp >= 65 & orings2$temp <70 ] <- 65
orings2$temp.grp[orings2$temp >= 70 & orings2$temp <75 ] <- 70
orings2$temp.grp[orings2$temp>=75] <- 75

orings3<-summaryBy(damage ~ temp.grp, data = orings2,
          FUN = function(x) { c(sum = sum(x), cnt=length(x)) } )
orings3$p<-orings3$damage.sum/(orings3$damage.cnt*6)
orings3$emp.logit<-log(orings3$p/(1-orings3$p))
orings3<-orings3[-3,]
```

Model Fit

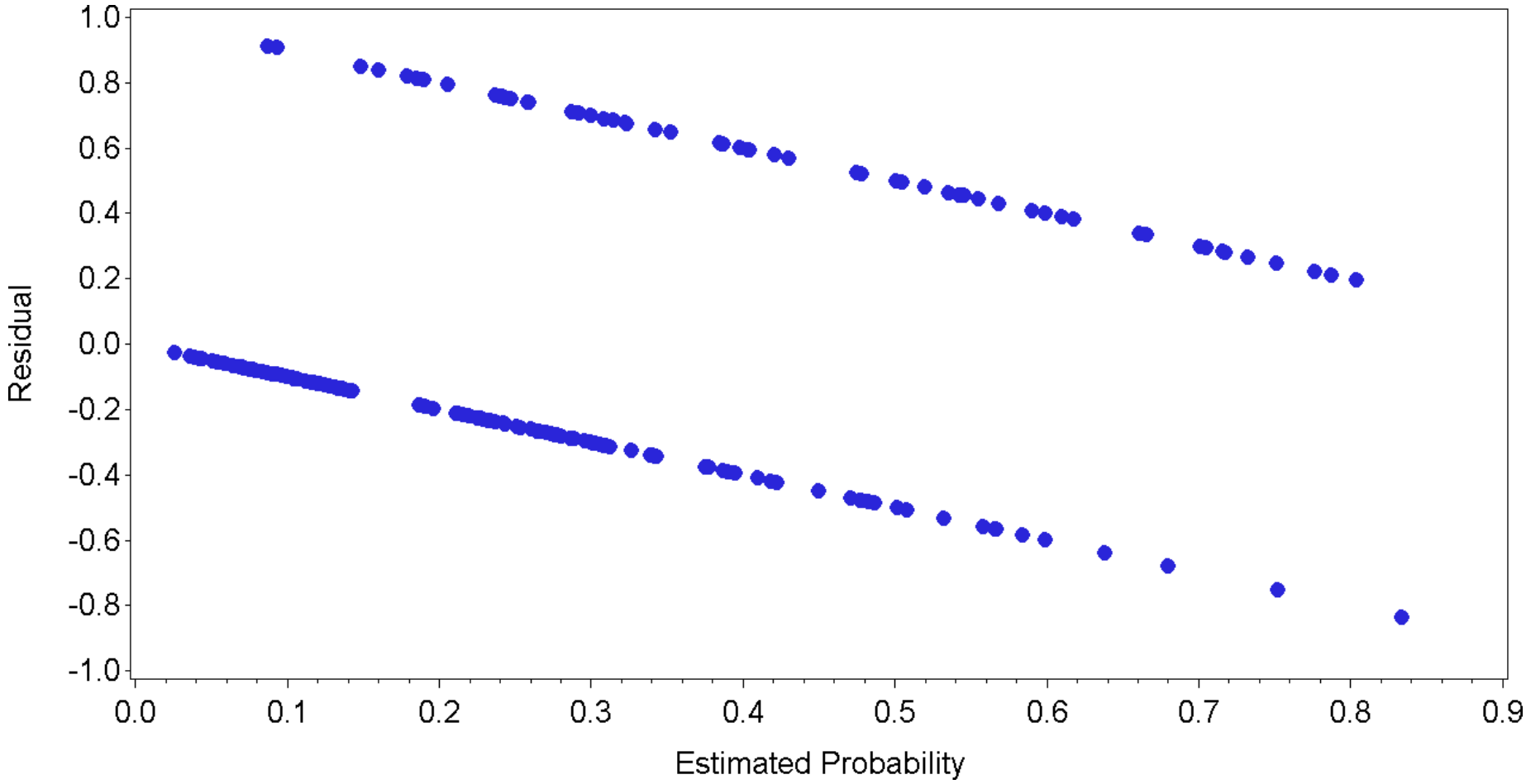
```
> orings3
  temp.grp damage.sum damage.cnt           p  emp.logit
1       55          7          3 0.38888889 -0.4519851
2       60          1          1 0.16666667 -1.6094379
4       70          2          6 0.05555556 -2.8332133
5       75          1          7 0.02380952 -3.7135721
```

Model Fit



Model Fit

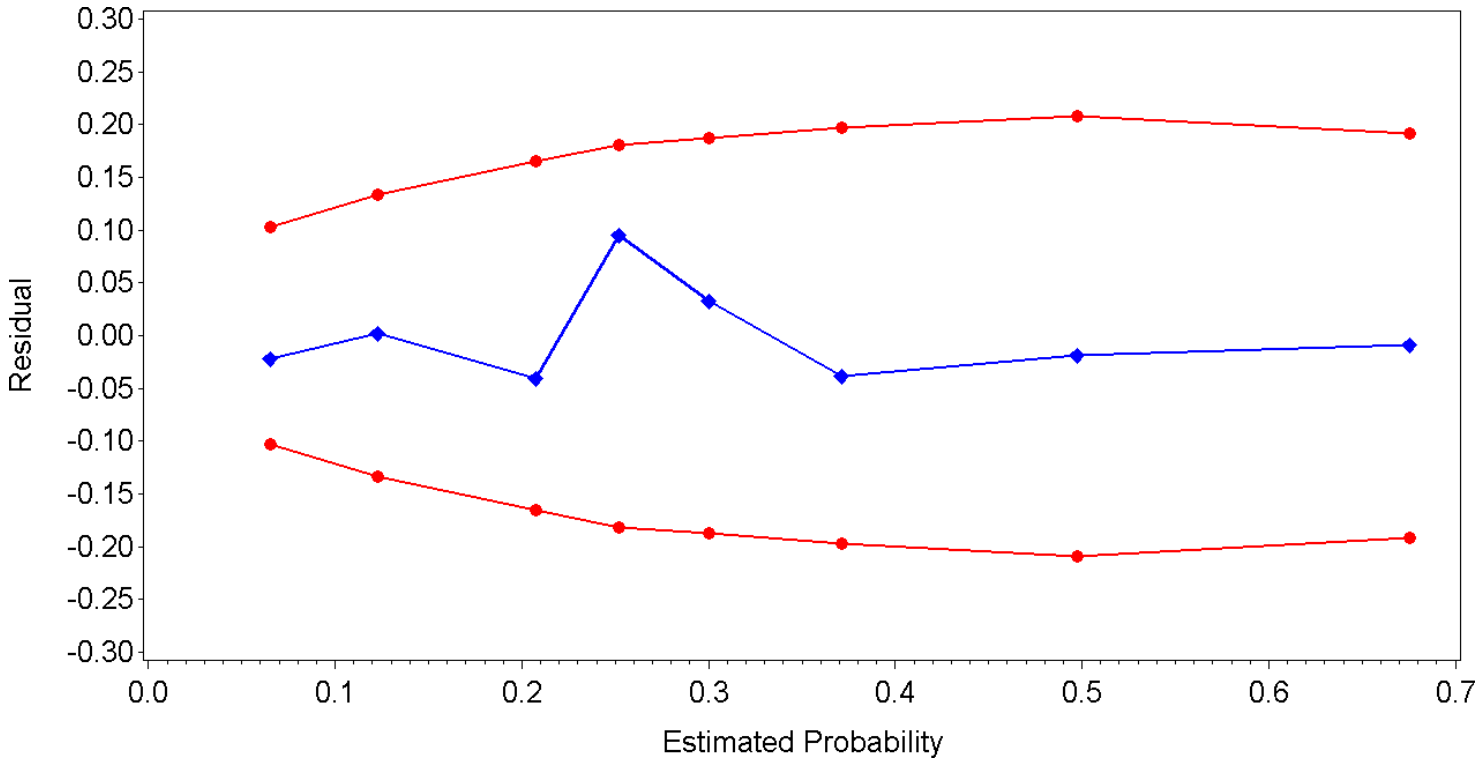
We can also look at the residuals. Here are the residuals plotted against the fitted values:



Model Fit

This is not very helpful as any given predicted value can only take on two values for the residual.

One method around this is to group cases, average their residuals and fitted values, and plot those. “Binned residual plot:”



Example 5

Table: Variables Available for Birth Weight Analysis

Variable	Description
ID	Identification Code
LOW	Low Birth Weight (0=Birth Weight greater than or equal to 2500g, 1=Birth Weight less than 2500g.)
AGE	Age of Mother in years
LWT	Weight in pounds at the last Menstrual Period
RACE	Race (1=White, 2=Black, 3=Other)
SMOKE	Smoking status During Pregnancy (1=Yes, 0=No)
PTL	History of Premature Labor (0=None, 1=One, etc.)
HT	History of Hypertension (1=Yes, 0=No)
UI	Presence of Uterine Irritability (1=Yes, 0=No)
FTV	Number of Physician Visits During the First Trimester (0=None, 1=One, 2=Two, etc.)
BWT	Birth Weight in grams.

Example 5

```
lowbwt.mod<-glm(LOW ~ AGE + LWT + factor(RACE) +
    SMOKE +    PTL + HT + UI + FTV,data=lowbwt)
```

Example 5

```
> lrtest(lowbwt.mod)
Likelihood ratio test

Model 1: LOW ~ AGE + LWT + factor(RACE) + SMOKE + PTL
Model 2: LOW ~ 1
#Df  LogLik Df Chisq Pr(>Chisq)
1  11 -105.89
2   2 -122.80 -9 33.81  9.643e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0
```

Example 5

```
> roc(factor(lowbwt$LOW),lowbwt$phat,plot=TRUE,auc.print=TRUE)
Setting levels: control = 0, case = 1
Setting direction: controls < cases

Call:
roc.default(response = factor(lowbwt$LOW), predictor = lowbwt$phat,      pl

Data: lowbwt$phat in 130 controls (factor(lowbwt$LOW) 0) < 59 cases (facto
Area under the curve: 0.7469
```

Example 5

```
> PseudoR2(lowbwt.mod,which="all")
McFadden     McFaddenAdj         CoxSnell      Nagelkerke
0.1376655       0.0562313       0.1638022       0.2252136
AldrichNelson VeallZimmermann           Efron McKelveyZavoina
0.1517445       0.2685201       0.1638022                NA
Tjur              AIC               BIC           logLik
NA     233.7867436     269.4459608    -105.8933718
logLik0              G2
-122.7984909      33.8102383
```

These are various Pseudo-R^2. It might be thought of as a standardized measure of improvement over the null model.

Not proportion of variance that is explained.

Example 5

We might say that the covariates add about 23% more information beyond knowledge of the mean.

Example 5

Let's reduce the model. See R.

Example 5

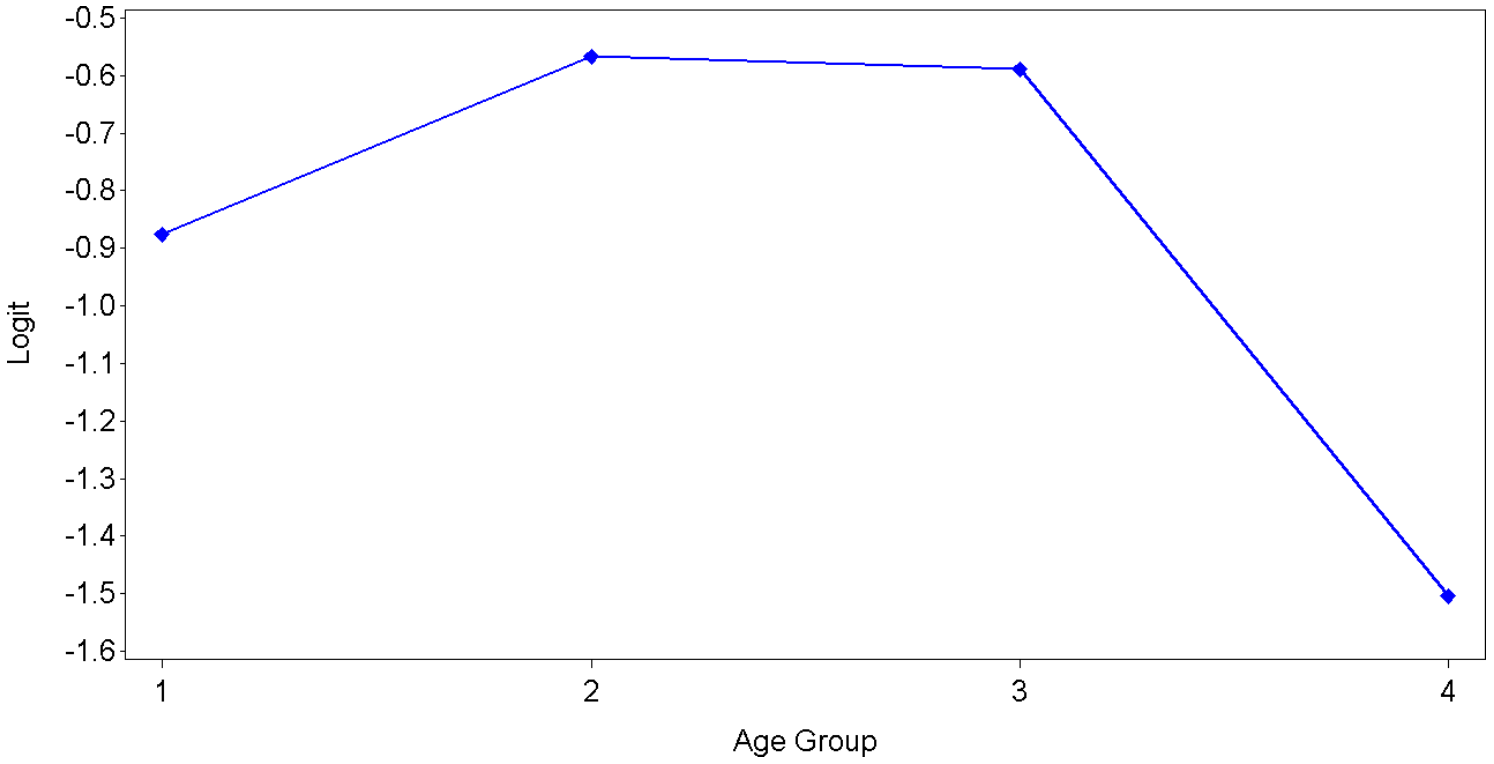
```
> Anova(lowbwt.mod, type="II", test="Wald")
Analysis of Deviance Table (Type II tests)

Response: LOW
               Df  Chisq Pr(>Chisq)
AGE             1 0.3296   0.565877
LWT             1 4.8142   0.028227 *
factor(RACE)    2 6.5050   0.038678 *
SMOKE           1 5.0295   0.024919 *
PTL             1 2.8770   0.089854 .
HT              1 7.3151   0.006838 **
UI              1 2.8347   0.092249 .
FTV             1 0.0413   0.838906
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Example 5

Is the relationship between AGE and LOW nonlinear?

Check the empirical logit plot:



Example 5

Try `AGE` and `AGE*AGE`:

```
> Anova(lowbwt.mod2, type="II", test="Wald")
Analysis of Deviance Table (Type II tests)

Response: LOW
           Df  Chisq Pr(>Chisq)
AGE         1 0.0280   0.867011
I(AGE^2)    1 0.0904   0.763636
LWT         1 3.4647   0.062692 .
RACE        1 3.9631   0.046509 *
SMOKE       1 5.1114   0.023769 *
PTL         1 2.8061   0.093906 .
HT          1 7.2650   0.007031 **
UI          1 2.7043   0.100078
FTV         1 0.0489   0.824986
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example 5

Reduce the model:

```
> Anova(lowbwt.mod3, type="II", test="Wald")
Analysis of Deviance Table (Type II tests)

Response: LOW
      Df  Chisq Pr(>Chisq)
LWT    1 4.7244   0.029738 *
RACE   1 5.0322   0.024881 *
SMOKE  1 7.3263   0.006795 **
HT     1 7.8663   0.005036 **
UI     1 4.3101   0.037887 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Example 5

Table: Comparing Two Models

Statistic	Model 2	Model 3
Log(L)	-107.110	-108.869
AIC	236.220	231.737
N-Pseudo R^2	0.210	0.188
AUC	0.746	0.728
p	9+Intercept	5+Intercept

Example 5

We can apply a test to the difference in the likelihoods, which will have a x^2 distribution.

$G_6 - G_{10} = -2(-107.110 - -105.893) = 2.434$ with $9 - 5 = 4$ df. This is not significant at the 0.01 level since $x^2_{4,.01} = 13.2$.

Example 5

Same type of test can be used to test interactions in the model.

```
>lowbwt.mod3<-glm(LOW ~ LWT + RACE + SMOKE + HT + UI,data=lowbwt)
>lowbwt.mod4<-glm(LOW ~ LWT + RACE + SMOKE + HT + UI + SMOKE*LWT,
 data=lowbwt)
> lrtest(lowbwt.mod4,lowbwt.mod3)
Likelihood ratio test

Model 1: LOW ~ LWT + RACE + SMOKE + HT + UI + SMOKE * LWT
Model 2: LOW ~ LWT + RACE + SMOKE + HT + UI
#Df  LogLik Df  Chisq Pr(>Chisq)
1   8 -108.81
2   7 -108.87 -1 0.1202     0.7288
```