

Categorical: Logistic regression 1

Table of contents

1	Goals	1
1.1	Goals	1
2	Review: Generalized linear model	2
2.1	GLiM	2
3	Logistic regression	3
3.1	Logistic regression	3
3.2	Binomial distribution	3
3.3	Model	8
3.4	Probability metric	13
3.5	Odds metric	18
3.6	Logit or log-odds metric	23
3.7	Metrics wrap-up	26
4	Summary	27
4.1	Summary	27

1 Goals

1.1 Goals

1.1.1 Goals of this lecture

- Introduce logistic regression
 - Binary outcome: Two mutually exclusive categories
 - * yes / no, pass / fail, diagnosed / not
 - Binomial distribution (vs normal)

- Logistic regression model
- Three (3!) metrics for interpretation

2 Review: Generalized linear model

2.1 GLiM

2.1.1 GLiM

- The generalized linear model (GLiM) is a not just one model
 - It is a **family** of regression models
 - Choose features (i.e., distribution) to match the characteristics of your outcome variable

2.1.2 Three components

- **Random component**
 - Distribution of the *outcome*
 - Exponential family: Normal, binomial, multinomial, Poisson
- **Systematic component**
 - *Linear combination* of predictors and regression coefficients
 - $\eta = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p$
- **Link function**
 - Relates (or *links*) random and systematic components
 - Transforms *predicted outcome* to link it to η

2.1.3 Some common GLiMs

Model	Outcome	Random	Link
Linear (OLS)	continuous, normal	normal	identity
Logistic	binary	binomial	logit
Probit	binary	binomial	probit
Ordinal logistic	ordered category	binomial	logit
Multinomial	unordered category	multinomial	logit
Poisson	count	Poisson	ln
Negative binomial	count	negative binomial	ln

3 Logistic regression

3.1 Logistic regression

3.1.1 (Binary) logistic regression

- Logistic regression is for **binary** outcomes
 - Exactly *two* categories, typically coded 0 and 1
 - * Typically define 1 as “success” or “event”
 - Predict “successes” or “events”
 - * Diagnosis, leave job, coin shows head
 - Predicted value is a **probability** (between 0 and 1)
 - Each observation is an independent **trial**

3.1.2 (Binary) logistic regression as GLiM

- **Outcome:** binary
 - Observed value (Y): 0 or 1, where 1 = “success” or “event”
 - Predicted value (\hat{Y}): **Probability** of success, between 0 and 1
- **Random component:** binomial
- **Link function:** logit (or log-odds) $= \ln\left(\frac{\hat{Y}}{1-\hat{Y}}\right)$

$$\ln\left(\frac{\hat{Y}}{1-\hat{Y}}\right) = \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p$$

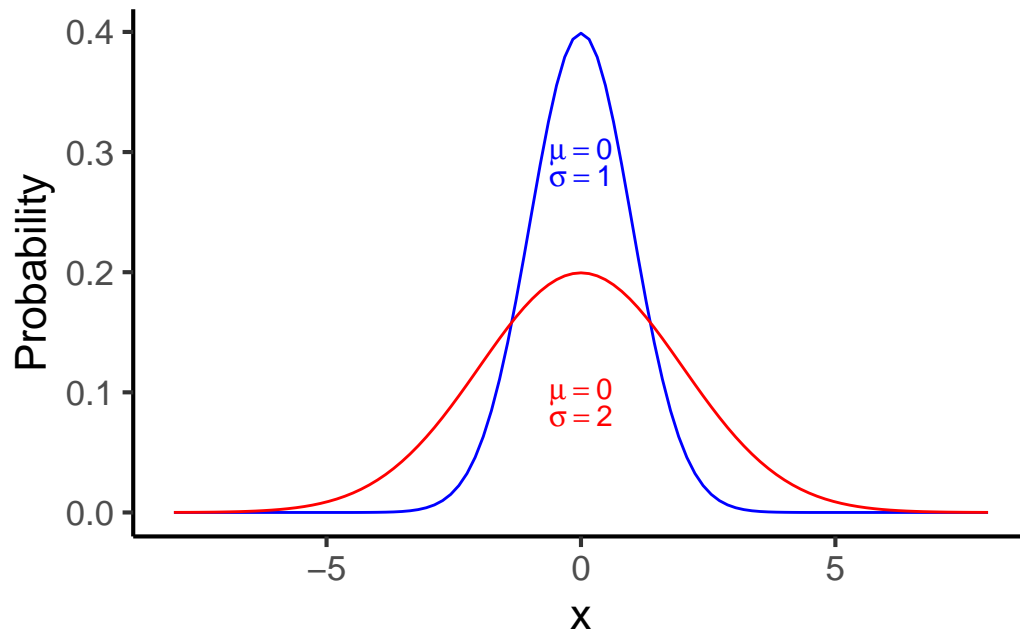
3.2 Binomial distribution

3.2.1 Actually, let's talk about the normal distribution

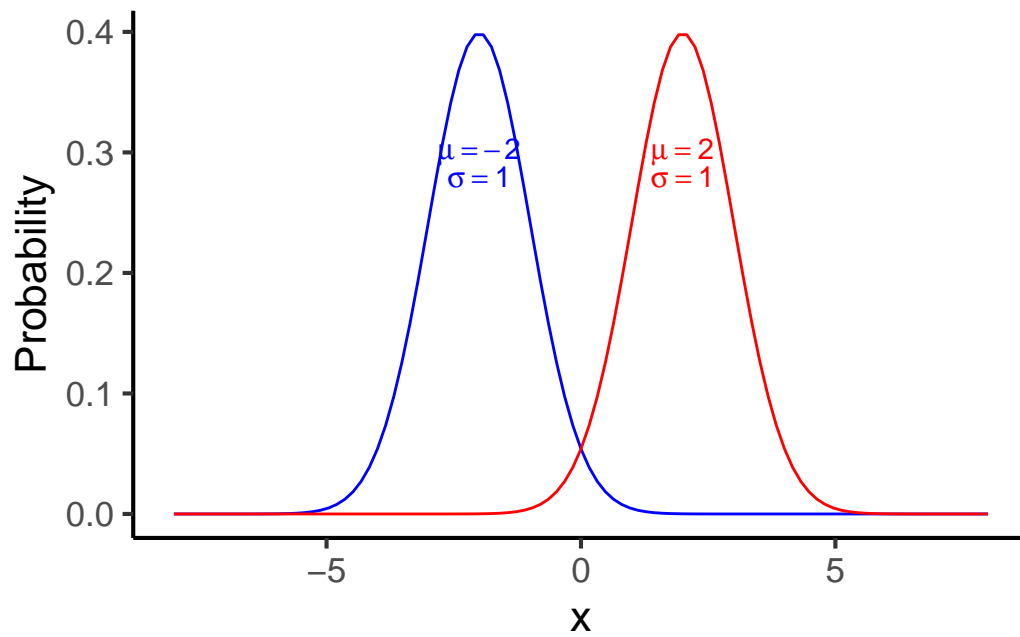
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mean $= \mu$
 - Location
- Variance $= \sigma^2$
 - Scale

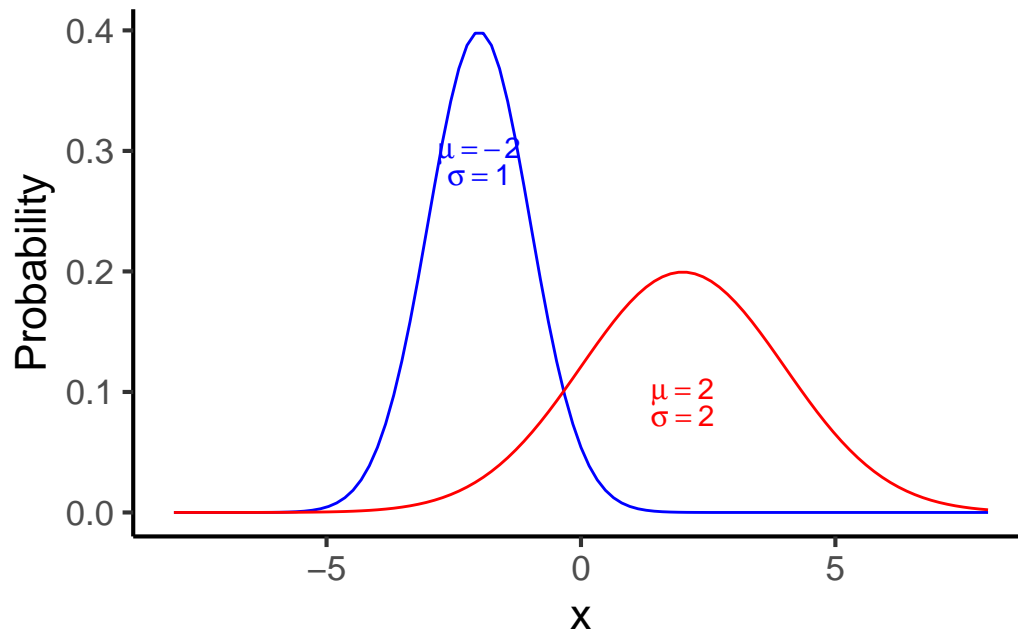
3.2.2 Same mean, different variance



3.2.3 Different mean, same variance



3.2.4 Different mean, different variance



3.2.5 Normal distribution

- Mean = μ
- Variance = σ^2
- Mean and variance are **independent**
 - Knowing one doesn't tell you about the other

3.2.6 (Binary) logistic regression

- Exactly *two* categories: 1 = “event”, 0 = “no event”
- Some number of independent *observations* (“trials”)
- Each observation has some *probability* of an event

3.2.7 Binomial distribution

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- Probability of *exactly* k events in n *independent* trials, each with probability p of an event

- Exactly $k = 5$ people in $n = 100$ are diagnosed if each person has a $p = 12\%$ chance of diagnosis?
- Exactly $k = 15$ heads in $n = 25$ coin flips if each coin has a $p = 50\%$ chance of heads?

3.2.8 Binomial distribution

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

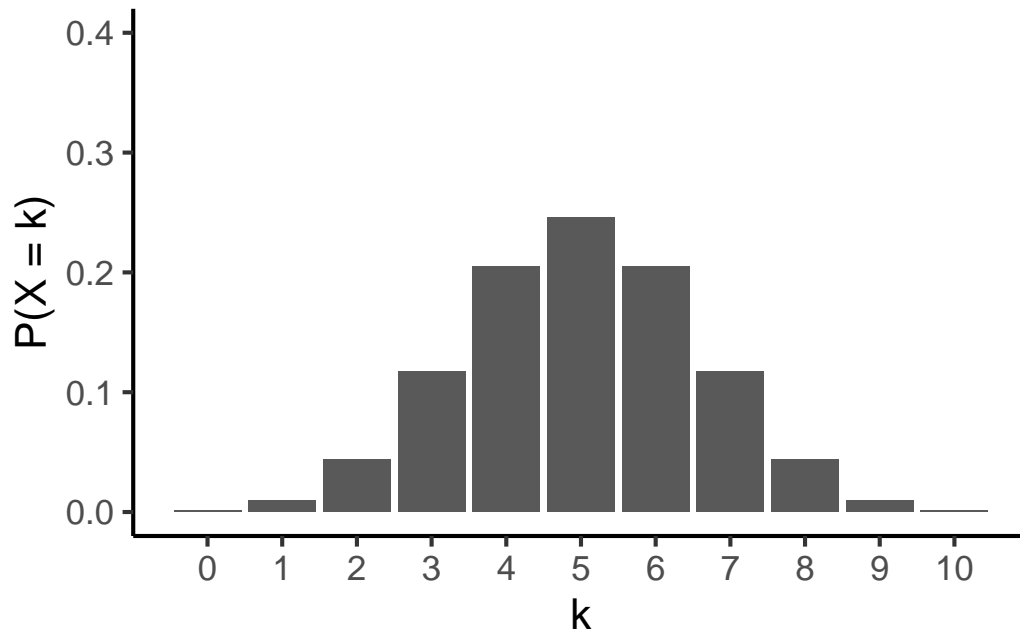
- n is the sample size
- k is the observed number of “events” or “successes”
- p is the probability of an “event” or “success”
- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ and is read as “ n choose k ”
 - It is the number of ways you can have k events in n trials

3.2.9 Binomial distribution

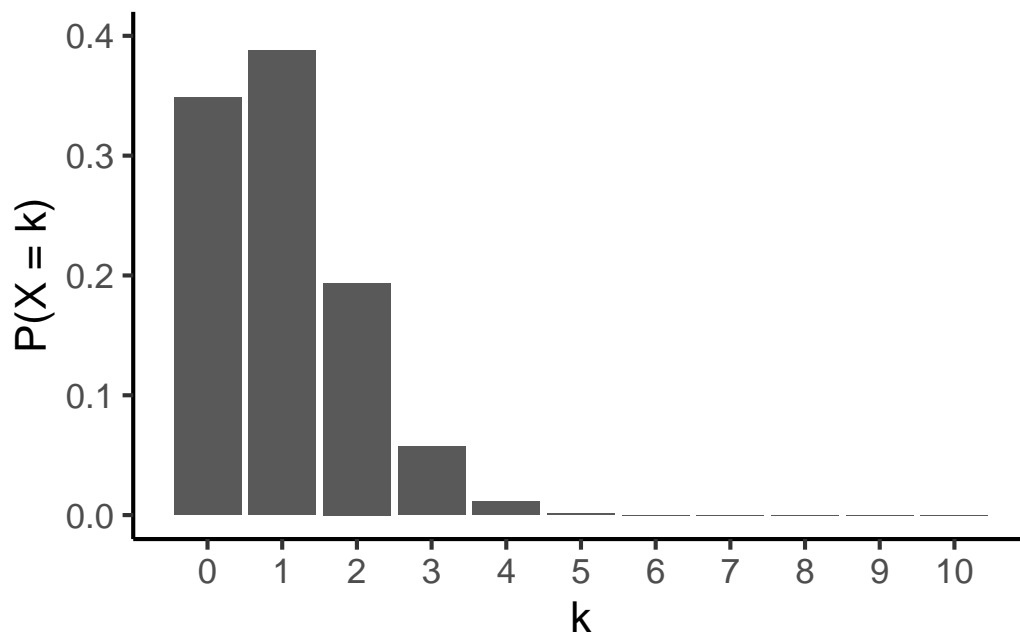
$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- Mean = np
- Variance = $np(1 - p)$
- Mean and variance are **not independent**
 - They are functions of the *same parameters*
 - Knowing about one gives you information about the other

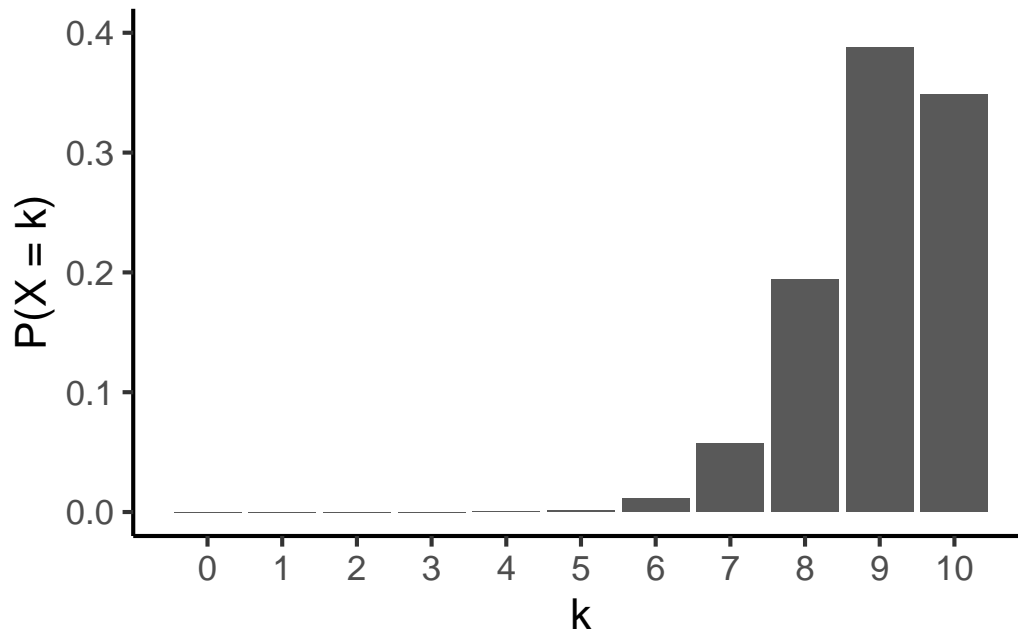
3.2.10 Binomial distribution: $n = 10, p = 0.5$



3.2.11 Binomial distribution: $n = 10, p = 0.1$



3.2.12 Binomial distribution: $n = 10, p = 0.9$



3.2.13 Logistic regression

- Linear regression models
 - **Mean** of the outcome, conditional on predictors(s)
 - Mean = μ and variance = σ^2
- Logistic regression models
 - **Probability** of a “success” or “event”, conditional on predictor(s)
 - Probability = p
 - * Mean = np and variance = $np(1 - p)$
 - * **Heteroskedasticity**: Variance changes with the mean

3.3 Model

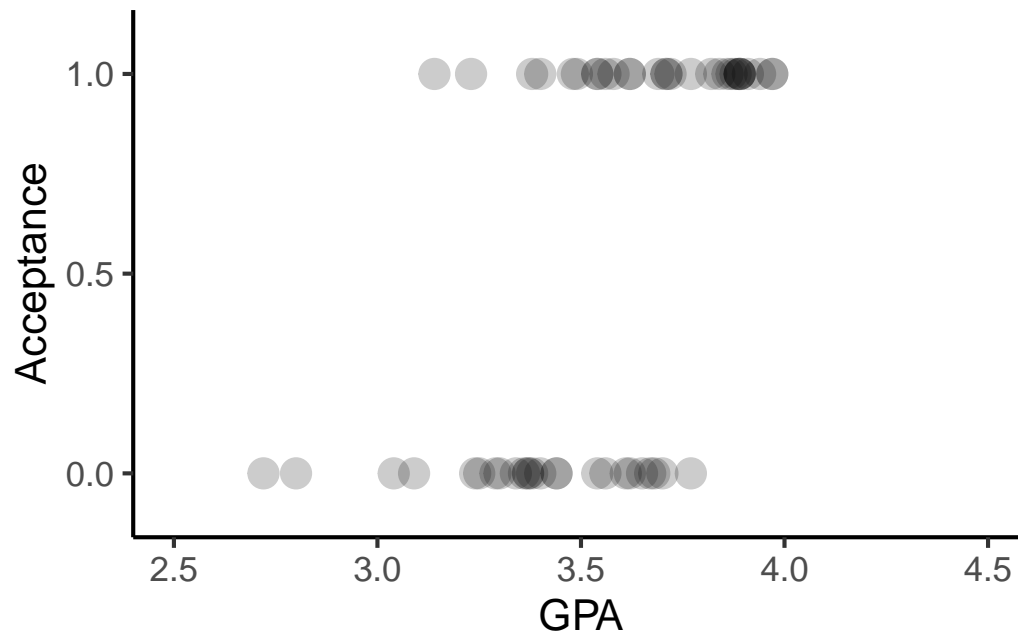
3.3.1 The data

- Stat2Data package: [info](#)
 - MedGPA dataset
 - * **Acceptance**: accepted in medical school (0 = no, 1 = yes)

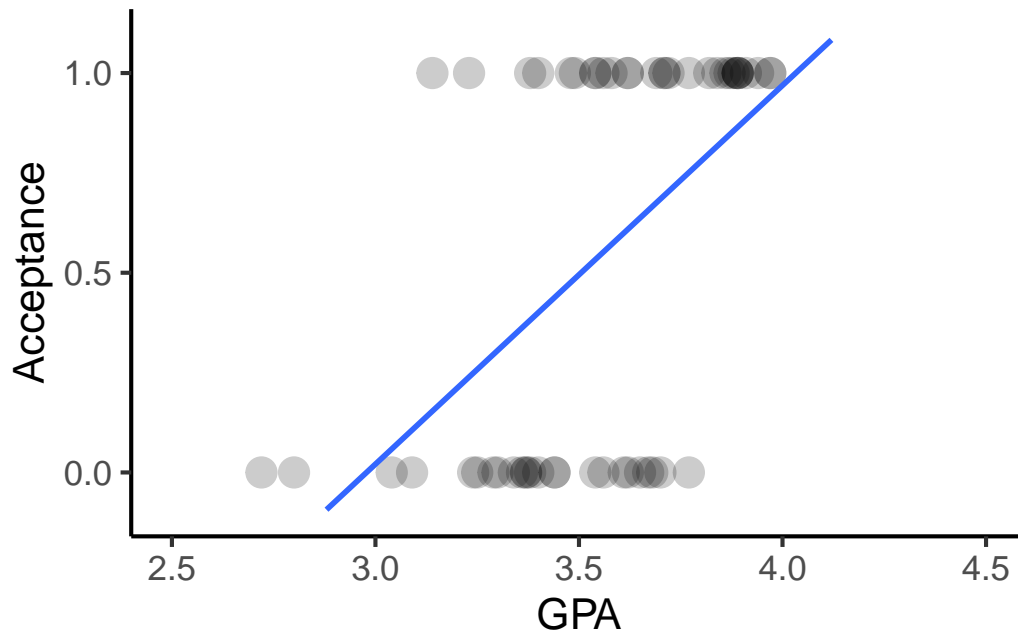
- * GPA: College GPA
- * MCAT: MCAT test score
- * Others

- Our model: GPA predicts Acceptance

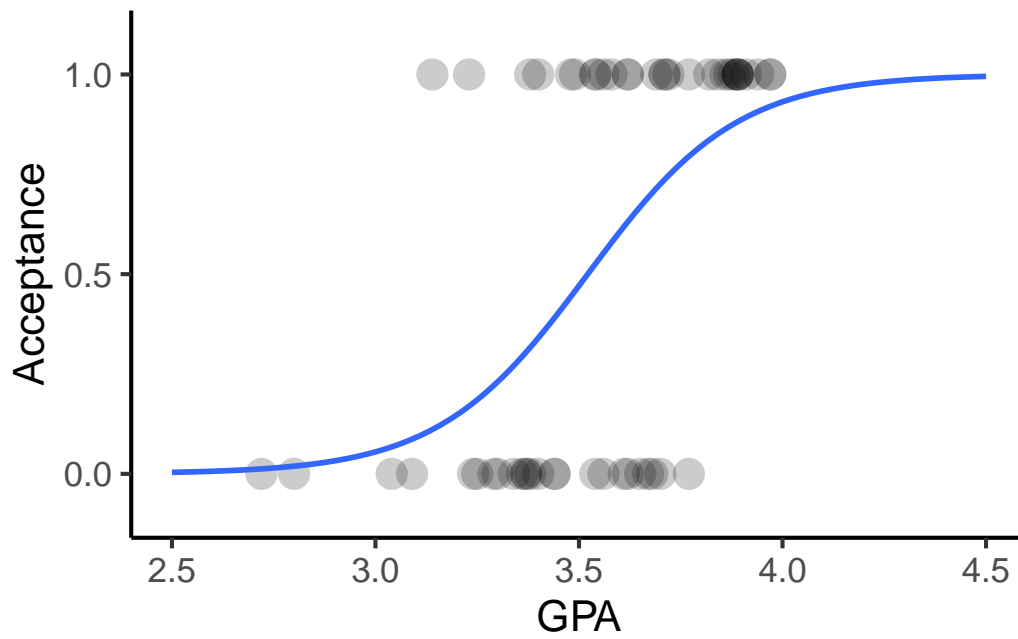
3.3.2 The data



3.3.3 The data, with a straight line



3.3.4 The data, with a not-straight line



3.3.5 Logistic regression model (mean centered GPA)

```
MedGPA <- MedGPA %>% mutate(GPAc = GPA - mean(GPA))
m1 <- glm(Acceptance ~ GPAc, MedGPA, family = binomial(link = "logit"))
summary(m1)
```

Call:

```
glm(formula = Acceptance ~ GPAc, family = binomial(link = "logit"),
     data = MedGPA)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7805	-0.8522	0.4407	0.7819	2.0967

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.1736	0.3253	0.534	0.593488
GPAc	5.4542	1.5792	3.454	0.000553 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 75.791 on 54 degrees of freedom
Residual deviance: 56.839 on 53 degrees of freedom
AIC: 60.839

Number of Fisher Scoring iterations: 4

3.3.6 Three forms of logistic regression

- **Observed** outcome: binary (1 = “success”, 0 = “not success”)
- **Predicted** outcome: Several options, all tell the same story
 - **Probability** of a “success” or “event”
 - * \hat{p} ranges from 0 to 1
 - **Odds** of a “success” or “event”
 - * $odds = \hat{p}/(1 - \hat{p})$
 - **Logit or log-odds** of a “success” or “event”
 - * $logit = \ln(odds)$

3.3.7 Three forms of logistic regression

Probability:

$$\hat{p} = \frac{e^{(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p)}}{1 + e^{(b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p)}}$$

Odds:

$$odds = \frac{\hat{p}}{1 - \hat{p}} = e^{b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p}$$

Logit:

$$\ln\left(\frac{\hat{p}}{1 - \hat{p}}\right) = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p$$

3.3.8 Three forms of logistic regression: Example

Probability:

$$\hat{p} = \frac{e^{(0.17 + 5.45(GPAc))}}{1 + e^{(0.17 + 5.45(GPAc))}}$$

Odds:

$$odds = \frac{\hat{p}}{1 - \hat{p}} = e^{0.17 + 5.45(GPAc)}$$

Logit:

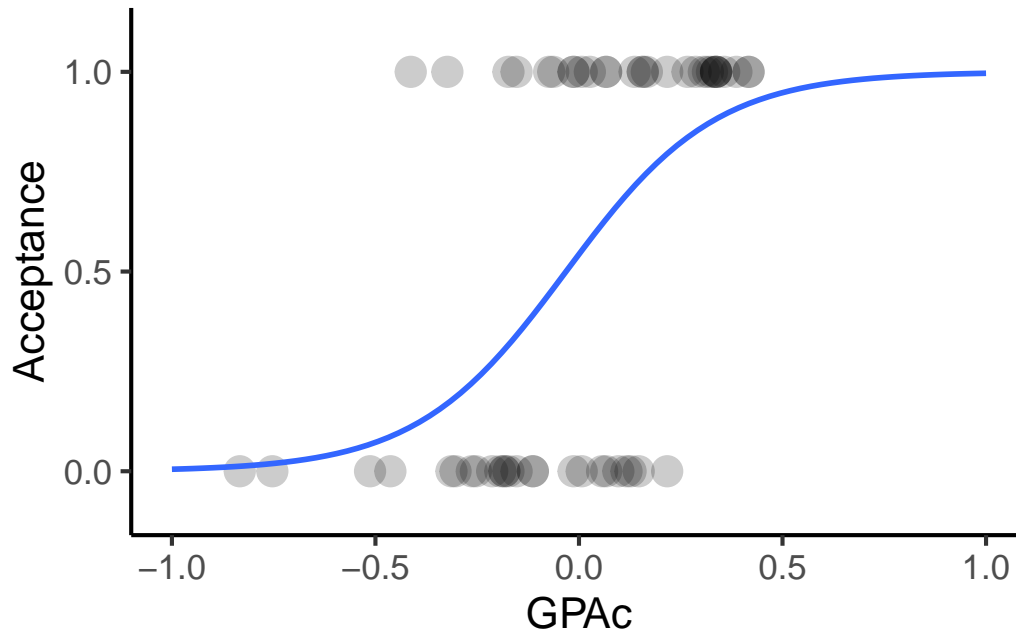
$$\ln\left(\frac{\hat{p}}{1 - \hat{p}}\right) = 0.17 + 5.45(GPAc)$$

3.4 Probability metric

3.4.1 What is probability (p)?

- Likelihood of a “success” or “event”
- Ranges from 0 to 1
- Both options are equally likely when $p = 0.5$

3.4.2 $\hat{p} = \frac{e^{0.17+5.45(GPAC)}}{1+e^{0.17+5.45(GPAC)}}$



3.4.3 Probability metric interpretation: General

$$\hat{p} = \frac{e^{0.17+5.45(GPAC)}}{1 + e^{0.17+5.45(GPAC)}}$$

- General interpretation of **intercept**:
 - b_0 is related to the **probability of success** when $\mathbf{X} = \mathbf{0}$
 - * $b_0 > 0$: Success (1) more likely than failure (0) when $\mathbf{X} = \mathbf{0}$
 - * $b_0 < 0$: Failure (0) more likely than success (1) when $\mathbf{X} = \mathbf{0}$

3.4.4 Probability metric interpretation: General

$$\hat{p} = \frac{e^{0.17+5.45(GPAc)}}{1 + e^{0.17+5.45(GPAc)}}$$

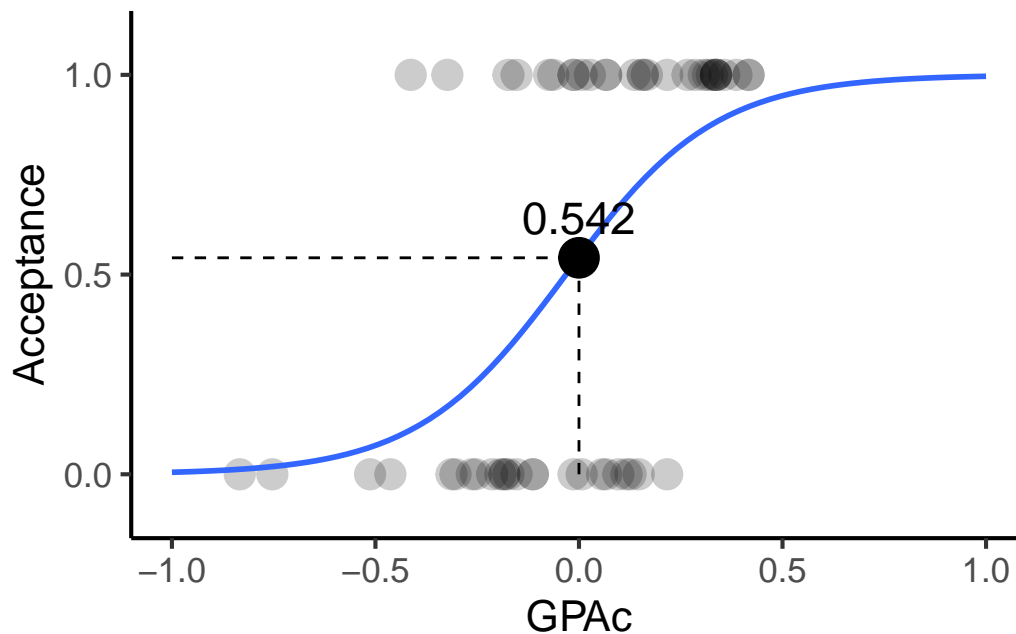
- General interpretation of **slope**:
 - b_1 tells you how **predictor X relates to probability of success**
 - * $b_1 > 0$: Probability of a success increases as X increases
 - * $b_1 < 0$: Probability of a success decreases as X increases

3.4.5 Probability metric interpretation: Example

$$\hat{p} = \frac{e^{0.17+5.45(GPAc)}}{1 + e^{0.17+5.45(GPAc)}}$$

- Interpretation of example **intercept**:
 - $b_0 > 0$: **Success (1) more likely than failure (0) when X = 0**
 - Probability of success when X = 0: $\frac{e^{b_0}}{1+e^{b_0}} = \frac{e^{0.17}}{1+e^{0.17}} = 0.542$

3.4.6 P(success|GPAc=0)



3.4.7 Probability metric interpretation: Example

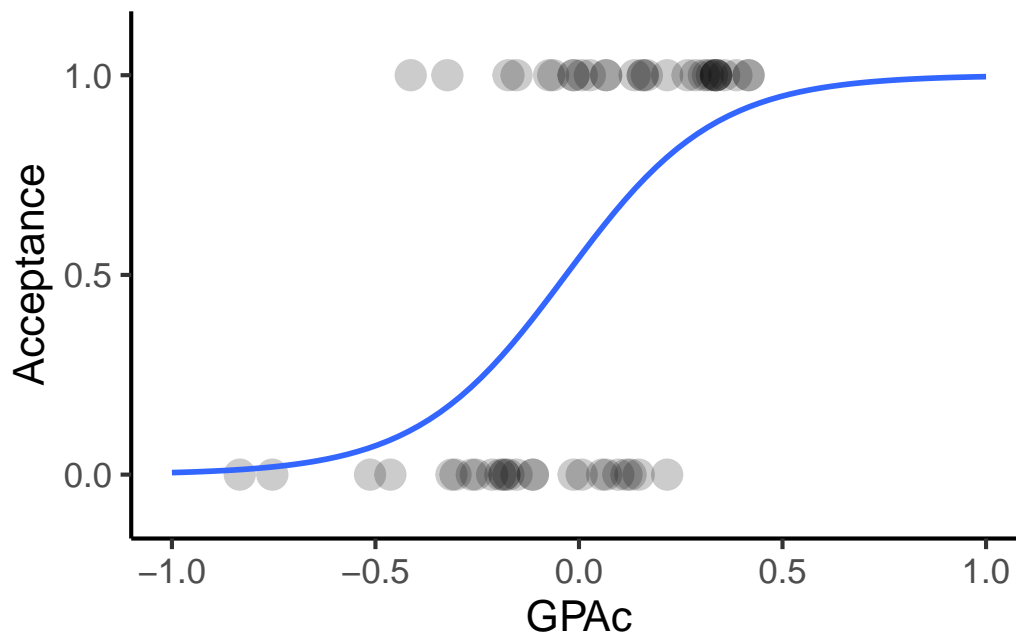
$$\hat{p} = \frac{e^{0.17+5.45(GPAc)}}{1 + e^{0.17+5.45(GPAc)}}$$

- Interpretation of example **slope**:
 - $b_1 > 0$: **Probability of a success increases as X increases**

3.4.8 Probability metric interpretation: Non-linear

- Linear regression:
 - **Constant**, linear slope
 - Slope depends on the **slope only**
- Logistic regression (probability):
 - **Non-linear** slope
 - Slope depends on BOTH **slope** (b_1) and **predicted probability** (\hat{p})
 - * The slope of the **tangent to the regression line** at the predicted outcome value = $\hat{p}(1 - \hat{p})b_1$

3.4.9 $\hat{p} = \frac{e^{0.17+5.45(GPAc)}}{1+e^{0.17+5.45(GPAc)}}$



3.4.10 Probability metric interpretation: Non-linear

When $GPAc = -0.25$:

$$\hat{p} = \frac{e^{b_0 + b_1 GPAc}}{1 + e^{b_0 + b_1 GPAc}} = \frac{e^{0.17 + 5.45 \times (-0.25)}}{1 + e^{0.17 + 5.45 \times (-0.25)}} = 0.233$$

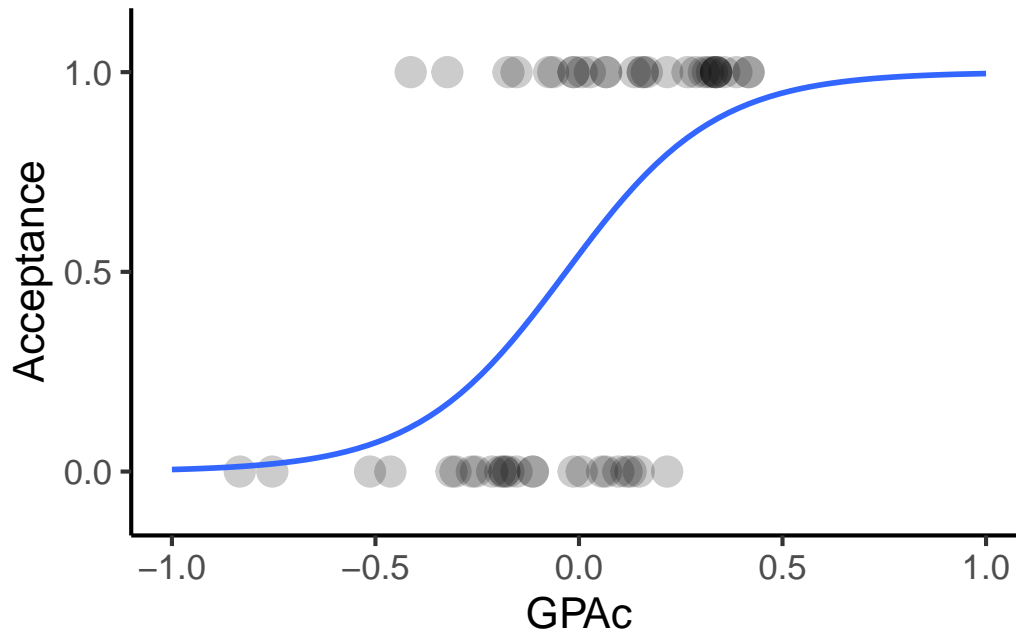
Approximate **slope** at that point is

$$\hat{p}(1 - \hat{p})b_1 = 0.233 \times (1 - 0.233) \times 5.45 = 0.974$$

3.4.11 Probability metric interpretation: Non-linear

X value	Predicted probability	Slope
-1.00	0.01	0.03
-0.75	0.02	0.10
-0.50	0.07	0.36
-0.25	0.23	0.97
0.00	0.54	1.35
0.25	0.82	0.80
0.50	0.95	0.27

$$3.4.12 \quad \hat{p} = \frac{e^{0.17+5.45(GPAC)}}{1+e^{0.17+5.45(GPAC)}}$$



3.4.13 A caution about probability equation

⚠ Warning

You might also see the probability defined as $\hat{p} = \frac{1}{1+e^{-(b_0+b_1X)}}$

Or more generally, $\hat{p} = \frac{1}{1+e^{-(Xb)}}$

- These are **numerically equivalent** to what we've talked about
 - But did you notice the *negative sign*?
 - No??? You didn't expect it and *missed it* in the complicated equation?
 - **Yeah, that's why we don't use this version:** No one likes a hiding negative sign...

3.4.14 Probability: Summary

- Probability is easy to understand and intuitive
- Non-linear relationship is harder to interpret

3.5 Odds metric

3.5.1 What are odds?

- *Odds* is the **ratio of two probabilities**
 - Model the probability of a “success”
 - Odds is the ratio of probability of a “success” (\hat{p}) to the probability of “not a success” ($1 - \hat{p}$)

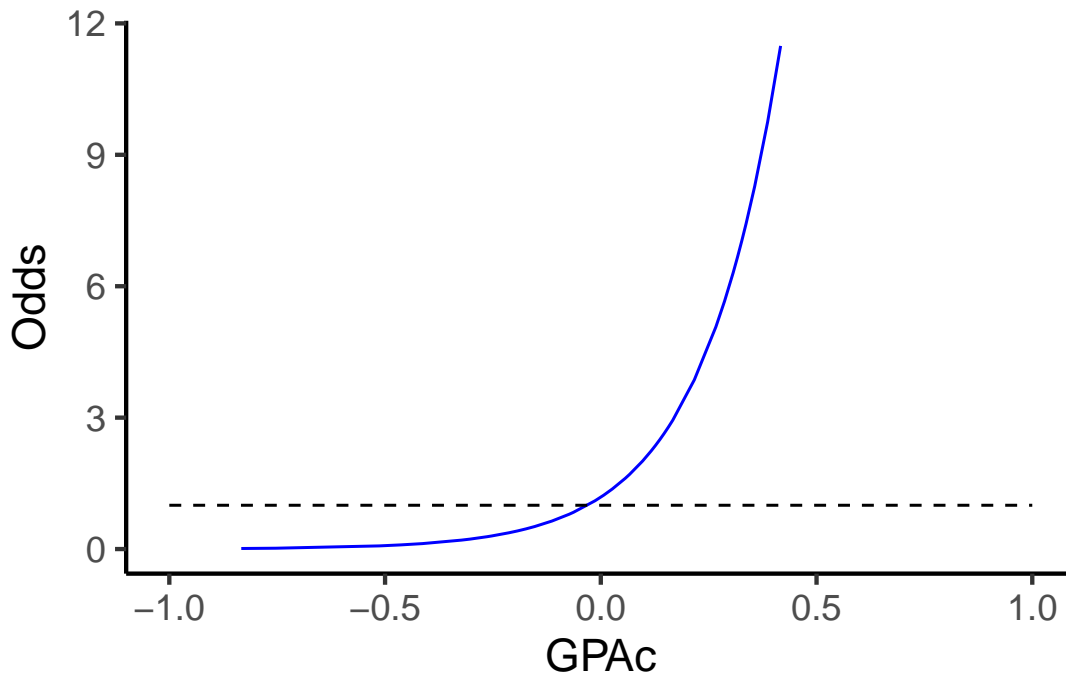
$$odds = \frac{\hat{p}}{(1 - \hat{p})}$$

- As **probability** of “success” increases (nonlinearly), the **odds** of “success” increases (also nonlinearly, **but in a different way**)

3.5.2 How do odds work?

- Probability ranges from 0 to 1, switches at 0.5
 - **Success more likely** than failure when $p > 0.5$
 - **Success less likely** than failure when $p < 0.5$
- Odds range from 0 to $+\infty$, switches at 1
 - **Success more likely** than failure when $odds > 1$
 - **Success less likely** than failure when $odds < 1$

3.5.3 $odds = \frac{\hat{p}}{(1-\hat{p})} = e^{0.17+5.45X}$



3.5.4 Odds metric interpretation: General

$$odds = \frac{\hat{p}}{(1-\hat{p})} = e^{0.17+5.45X}$$

- General interpretation of **intercept**:
 - b_0 is related to the **odds of success when $X = 0$**
 - * Odds of success **when $X = 0$** : e^{b_0}
 - * $b_0 > 0$: Odds of success > 1 when $X = 0$
 - * $b_0 < 0$: Odds of success < 1 when $X = 0$

3.5.5 Odds metric interpretation: General

$$odds = \frac{\hat{p}}{(1-\hat{p})} = e^{0.17+5.45X}$$

- General interpretation of **slope**:
 - b_1 = relationship between predictor X and the odds of success

- * $b_1 > 0$: Odds of success **increases** as X **increases**
- * $b_1 < 0$: Odds of a success **decreases** as X **increases**

3.5.6 Odds metric interpretation: Example

$$odds = \frac{\hat{p}}{(1 - \hat{p})} = e^{0.17 + 5.45X}$$

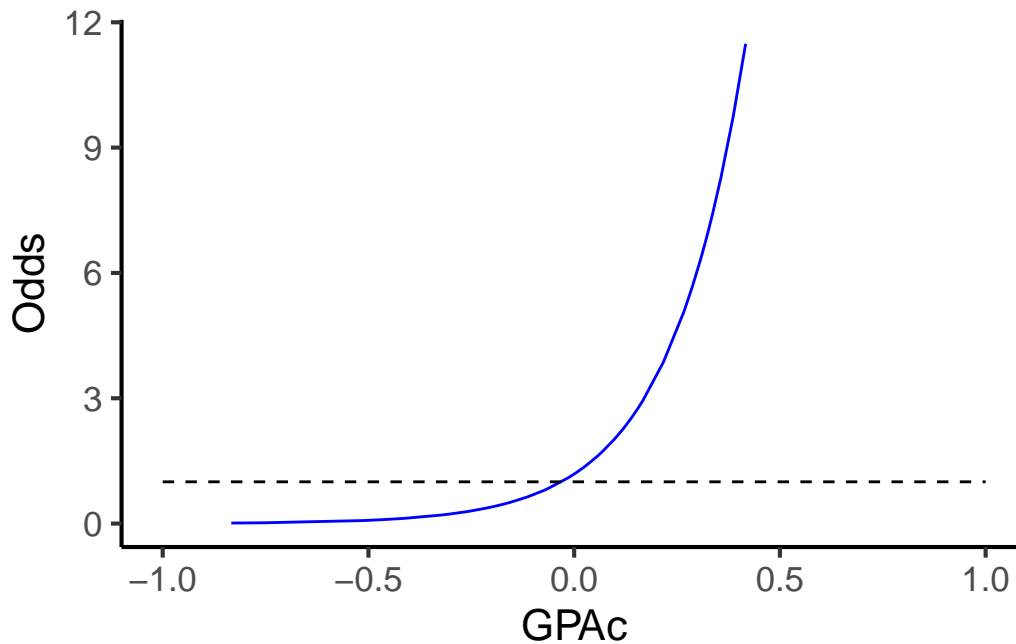
- Interpretation of example **intercept**:
 - $b_0 > 0$: Odds of success > 1 when $X = 0$
 - * Success (1) more likely than failure (0) when $X = 0$
 - * Odds of success when $X = 0$: $e^{b_0} = e^{0.17} = 1.19$
 - * A “success” is about 1.19 times as likely as a “failure”
 - * Compare to $\hat{p} = 0.542$: $0.542 / 0.458 = 1.18$

3.5.7 Odds metric interpretation: Example

$$odds = \frac{\hat{p}}{(1 - \hat{p})} = e^{0.17 + 5.45X}$$

- Interpretation of example **slope**:
 - $b_1 > 0$: Odds of a success **increases** as X **increases**

3.5.8 Odds metric interpretation: Non-linear



3.5.9 Odds metric interpretation: Non-linear

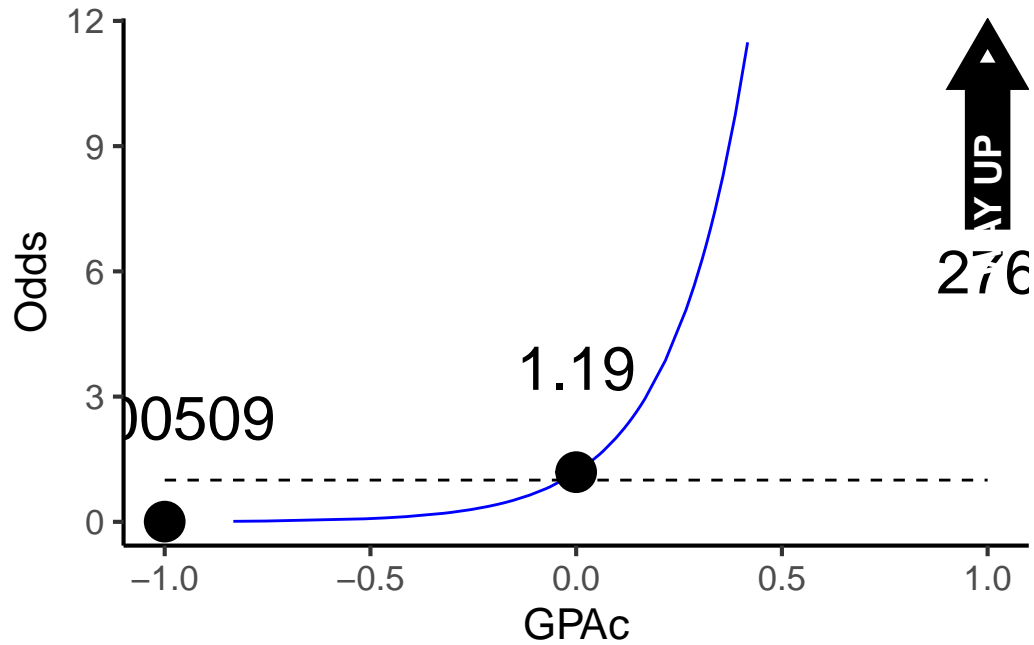
- This non-linear change is presented in terms of **odds ratio**
 - Constant, **multiplicative** change in predicted odds
 - For a 1-unit difference in X , the **predicted odds** of success is **multiplied by the odds ratio**
- *Example:* odds ratio $= e^{b_1} = e^{5.45} = 232.8$
 - For a 1-unit difference in X , the **predicted odds** of success is **multiplied by 232.8**

3.5.10 Odds metric interpretation: Non-linear

- Odds ratio $= e^{b_1} = e^{5.45} = 232.8$
- Odds ratio for $X = 0$ versus $X = -1$: $\frac{odds(X=0)}{odds(X=-1)} = \frac{1.1853049}{0.0050924} = 232.8$
 - Odds of success is 232.8 times larger when $X = 0$ vs $X = -1$
- Odds ratio for $X = 1$ versus $X = 0$: $\frac{odds(X=1)}{odds(X=0)} = \frac{275.8893832}{1.1853049} = 232.8$
 - Odds of success is 232.8 times larger when $X = 1$ vs $X = 0$

- Any 1 unit difference in X : **Constant multiplicative change**

3.5.11 Odds metric figure again (odds ratio = 232.8)



3.5.12 Odds metric interpretation: Non-linear

X value	Predicted probability	Predicted odds
-1.00	0.01	0.01
-0.75	0.02	0.02
-0.50	0.07	0.08
-0.25	0.23	0.30
0.00	0.54	1.19
0.25	0.82	4.63
0.50	0.95	18.08

3.5.13 A caution about odds

Warning

- **Odds ratios** are popular in medicine and epidemiology
- They can be **extremely** misleading
- The **same odds ratio** corresponds to many **different** probability values
 - Odds ratio = $\frac{odds=3}{odds=1} = 3$
 - * Corresponds to probability of 0.75 vs 0.5
 - Odds ratio = $\frac{odds=9}{odds=3} = 3$
 - * Corresponds to probability of 0.90 vs 0.75

3.5.14 Odds: Summary

- Odds can be difficult to understand (unless you gamble)
- Non-linear relationship is harder to interpret

3.6 Logit or log-odds metric

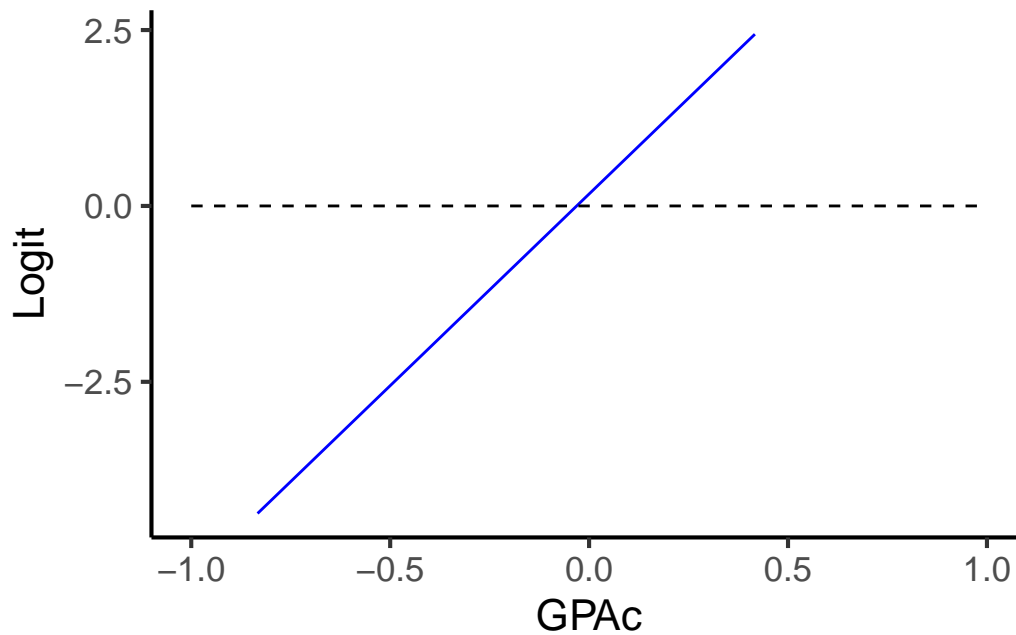
3.6.1 What is the logit?

- **Logit** or **log-odds** is the natural log (\ln) of the odds
 - As **probability** of “success” increases (nonlinearly, S-shaped curve)
 - * *Odds* of “success” increases (also nonlinearly, exponentially up)
 - * **Logit** of “success” increases **linearly**

3.6.2 How does the logit work?

- Probability ranges from 0 to 1, **switches at 0.5**
- Odds range from 0 to $+\infty$, **switches at 1**
- Logit ranges from $-\infty$ to $+\infty$, **switches at 0**
 - Success more likely than failure when logit > 0
 - Success less likely than failure when logit < 0

3.6.3 $\widehat{logit} = \ln\left(\frac{\hat{p}}{(1-\hat{p})}\right) = 0.17 + 5.45X$



3.6.4 Logit metric interpretation: General

$$\widehat{logit} = \ln\left(\frac{\hat{p}}{(1-\hat{p})}\right) = 0.17 + 5.45X$$

- General interpretation of **intercept**:
 - b_0 is related to the logit of success when $X = 0$
 - * Logit of success when $X = 0$: b_0
 - * $b_0 > 0$: Logit > 0 when $X = 0$
 - * $b_0 < 0$: Logit < 0 when $X = 0$

3.6.5 Logit metric interpretation: General

$$\widehat{logit} = \ln\left(\frac{\hat{p}}{(1-\hat{p})}\right) = 0.17 + 5.45X$$

- General interpretation of **slope**:
 - b_1 is the relationship between predictor X and logit of success
 - * $b_1 > 0$: Logit of a success increases as X increases
 - * $b_1 < 0$: Logit of a success decreases as X increases

3.6.6 Logit metric interpretation: Example

$$\text{logit} = \ln \left(\frac{\hat{p}}{(1 - \hat{p})} \right) = 0.17 + 5.45X$$

- Interpretation of example **intercept**
 - $b_0 > 0$: Logit > 0 when $X = 0$
 - Logit of success when $X = 0$: $b_0 = 0.17$

3.6.7 Logit metric interpretation: Example

$$\text{logit} = \ln \left(\frac{\hat{p}}{(1 - \hat{p})} \right) = 0.17 + 5.45X$$

- Interpretation of example **slope**
 - $b_1 > 0$: Logit of a success increases by 5.45 units when X increases by 1 unit

3.6.8 Logit metric interpretation: Linear!

X value	Predicted probability	Predicted odds	Predicted logit
-1.00	0.01	0.01	-5.28
-0.75	0.02	0.02	-3.92
-0.50	0.07	0.08	-2.56
-0.25	0.23	0.30	-1.19
0.00	0.54	1.19	0.17
0.25	0.82	4.63	1.53
0.50	0.95	18.08	2.90

3.6.9 Logit: Summary

- Hey, a linear relationship! Yay!
- But what's a logit?
 - Why do I care about a 5.45 unit change in it?

3.7 Metrics wrap-up

3.7.1 So which metric should I use?

- They are **equivalent**, so use the metric that
 - Makes the most *sense* to you
 - You can *explain* fully
 - Is most commonly used in your *field*

3.7.2 Some things to keep in mind

- Odds ratios tell you about **change**, but not **where you start**
 - If you report odds ratios, *also report some measure of probability* e.g., probability of success at the mean of X
 - 10x change is 5 to 50 or 0.05 to 0.5?
- Logit is nice because it's **linear**, but it's not very **interpretable**
 - What is a “logit”? It's just a *mathematical* concept that makes a straight line – not actually **meaningful**
 - *But many psychology measures don't have meaningful metrics...*

3.7.3 Example interpretation

- **Probability of Acceptance** increases with GPA
 - About 50/50 at mean GPA (~ 3.5)
 - Increase faster from GPA = 3.5 to 4.0 than from 2.5 to 3.0
- **Odds of Acceptance** increases with GPA
 - Constant *multiplicative* increase of 232.8 (odds ratio)
- **Logit of Acceptance** increases with GPA
 - Constant *linear* increase of 5.45

4 Summary

4.1 Summary

4.1.1 Summary of this week

- Normal distribution vs binomial distribution
- Logistic regression
 - Probability
 - Odds
 - Logit

4.1.2 Next week

- More logistic regression
 - Pseudo- R^2 measures
 - Tests of coefficients
 - Comparing models
 - * e.g., model with GPA vs model with GPA and MCAT

4.1.3 Next few weeks

- Extend logistic regression to 3 or more categories:
 - Ordinal logistic regression
 - Multinomial logistic regression
- Count outcomes: Poisson regression, overdispersed Poisson regression, negative binomial regression, excess zeroes versions of these models