

# Assessing Model Fit II

# Announcements

- ★ Notes for Nicky  
→ hispanic/latinx → why miss?  
→
- SPH Student Survey: ✓
    - [https://ohsu.ca1.qualtrics.com/jfe/form/SV\\_0OH1eqNZtawhhcy](https://ohsu.ca1.qualtrics.com/jfe/form/SV_0OH1eqNZtawhhcy)
  - Quarto workshop 6/10 @ 9:30 - 5:30
    - <https://sites.google.com/view/asa-oregon-chapter-short-cours>
    - Fee waived for first 25 students who sign up!
    - Attending half, and leaving after the lunch break could work?
  - Will post Midterm feedback form by end of this week
    - Should take about 10 minutes → decide on final pt value

# Class 15 Learning Objectives

1. Revisit Pearson Residual: How to extract Pearson residuals and other summary measures in R
2. Apply AIC and BIC as a summary measure to make additional comparisons between potential models
3. Understand the components of calculations for logistic regression diagnostics
4. <sup>A</sup>Plot and determine observations where regression does not fit well or are influential using specific diagnostic values

# Class 15 Learning Objectives

1. Revisit Pearson Residual: How to extract Pearson residuals and other summary measures in R
2. Apply AIC and BIC as a summary measure to make additional comparisons between potential models
3. Understand the components of calculations for logistic regression diagnostics
4. Plot and determine observations where regression does not fit well or are influential using specific diagnostic values

# Pearson Residual (I)

- In logistic regression model, can use Pearson residual for summary measure of goodness-of-fit
  - Uses the  $\hat{y}_j$  fitted value from previous slide

$$\hat{y}_j = m_j \hat{\pi}_j$$

- Pearson residual for jth covariate pattern is:

$$r(y_j, \hat{\pi}_j) = \frac{(y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} = \frac{(y_j - \hat{y}_j)}{\sqrt{\hat{y}_j (1 - \hat{\pi}_j)}}$$

$j$  = covariate pattern

- The summary statistics of Pearson residual is thus:

$$X^2 = \sum_{j=1}^J r(y_j, \hat{\pi}_j)^2$$

# Pearson residual: key info for testing

- Hypotheses:
  - $H_0$ : model fits data well
  - $H_1$ : model does not fit data well

- Test statistic for HL test

$$\rightarrow X^2 = \sum_{j=1}^J r(y_j, \hat{\pi}_j)^2$$

- HL test statistic follows a chi-squared distribution

$$X^2 \sim \chi^2_{df=J-(p+1)}$$

J is the number of covariate patterns and p is number of coefficients ( $\beta_1, \dots$ )

# Pearson Residuals: GLOW Study

- Let's say we decide the following is our **preliminary final model**:

$$\text{logit}(\pi(x)) = \beta_0 + \beta_1 I(PF_i = \text{"Yes"}) + \beta_2 I(MF_i = \text{"Yes"}) + \beta_3 I(AA_i = \text{"Yes"}) + \beta_4 I(RR = \text{"Greater"}) + \beta_5 I(MF_i = \text{"Yes"}) \times I(AA_i = \text{"Yes"})$$

- How many covariate patterns do we have?

$$(\# \text{ categories of } PF) \times (\# \text{ categories of } MF) \times (\# \text{ categories of } AA) \times (\# \text{ categories of } RR) = 2 \times 2 \times 2 \times 2 = 16$$

<sup>2<sup>4</sup></sup>  
~~2 × 2 × 2 × 3~~

- How many degrees of freedom do we have?

$$J - (p + 1) = 16 - (5 + 1) = 10$$

$\downarrow$   
 $p$

# Pearson Residuals: GLOW Study

- Fit the preliminary final model (categorical version)

```
cat_model = glm(frac_num ~ priorfrac + momfrac +
                 armassist + raterisk2 +
                 momfrac*armassist,
                 data = glow2, family = "binomial")
```

Logistic Dx

```
gof(cat_model)$chiSq
```

```
##      test      chiSq   df    pVal
## 1:  PrI  497.234325 494 0.4507076
## 2:  drI  523.939507 494 0.1697839
## 3: PfG  4.756822  10 0.9068193
## 4: drG  4.686232  10 0.9111289
## 5: PrCT 4.756822  10 0.9068193
## 6: drCT 4.686232  10 0.9111289
```

$H_0$ : model fits well  
 $H_1$ : model does not fit well

fail to reject null  
model fits data well

# Class 15 Learning Objectives

1. Revisit Pearson Residual: How to extract Pearson residuals and other summary measures in R
2. Apply AIC and BIC as a summary measure to make additional comparisons between potential models
3. Understand the components of calculations for logistic regression diagnostics
4. Plot and determine observations where regression does not fit well or are influential using specific diagnostic values

# AIC and BIC (I)

- Two widely used non-hypothesis testing based measurements that helps select a good model
  - Akaike Information Criterion (AIC)
  - Bayesian Information Criterion (BIC)
- Unlike likelihood ratio test which is only suitable for nested model, AIC and BIC are suitable for both nested and non-nested model
- There is no hypothesis/conclusion testing for the comparison between two models
  - So not the best for selecting covariates to include in model
  - BUT helpful if you have a few preliminary final models that you want to compare

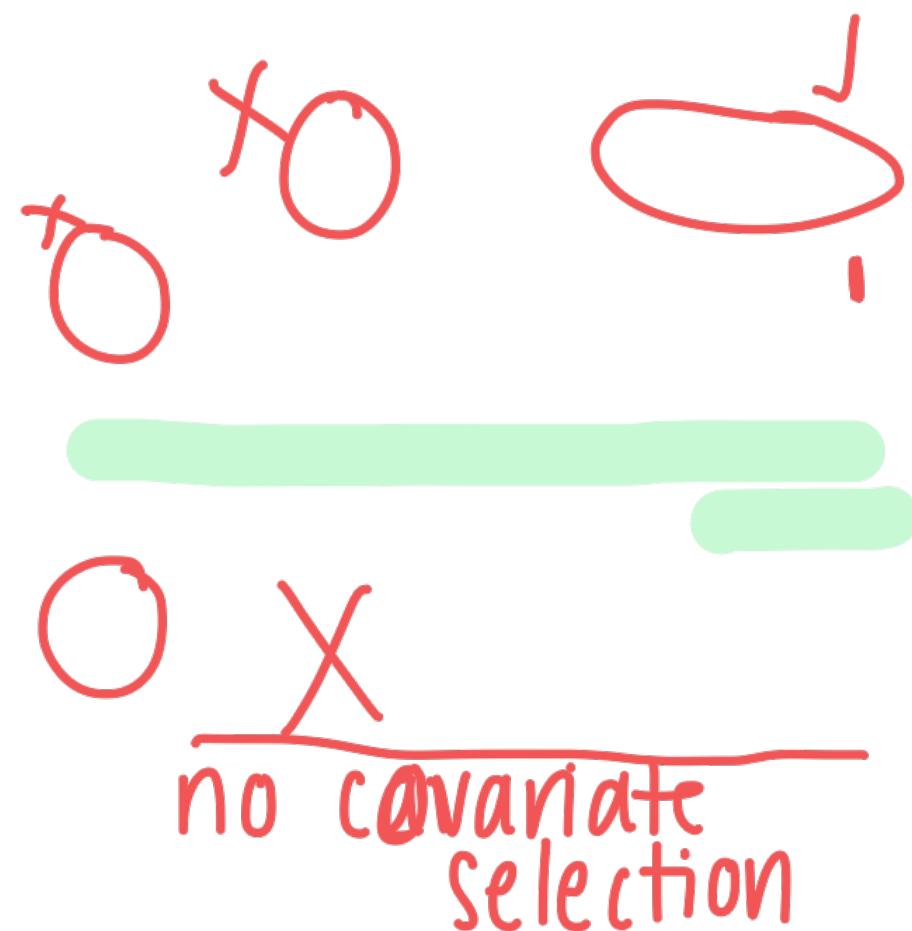
# Poll Everywhere

## Question 1

Can I compare these two models using AIC and BIC?  $\text{logit}(\pi(\text{age}_i)) = \beta_0 + \beta_1 \text{age}_i$  and  $\text{logit}(\pi(\text{height}_i)) = \beta_0 + \beta_1 \text{height}_i$



[https://www.polleverywhere.com/multiple\\_choice\\_polls/ven54AVKtq8d2Ad5MiWvz](https://www.polleverywhere.com/multiple_choice_polls/ven54AVKtq8d2Ad5MiWvz)



## AIC and BIC (II)

- Both AIC and BIC penalize a model for having many parameters

$$AIC = \text{circled } -2 \log likelihood + \text{circled } 2q$$

*aka covariates*

$$BIC = -2 \log likelihood + \text{circled } q \log(n)$$

Where  $q$  is the number of parameters in the model and  $n$  is the sample size

- Both AIC and BIC can **only** be used to compare models fitting the **same data set**

- In comparing two models, **the model with smaller AIC and/or BIC is preferred**

- When the difference in AIC between two models exceeds 3, the difference is viewed as “meaningful”



$$\begin{array}{ll} \text{AIC} = \cancel{3000} & 500 \\ \text{Model 1} & 497 \\ \text{AIC} = \cancel{2997} & \\ \text{Model 2} & \end{array}$$

# AIC and BIC in R and SAS

- After fitting the logistic regression model, can calculate AIC and BIC

```
prelim_final = glm(fracture ~ age + height + priorfrac +  
momfrac + armassist + raterisk2 +  
age*priorfrac + momfrac*armassist,  
data = glow2, family = binomial)
```

AIC(prelim\_final)

## [1] 518.4966

BIC(prelim\_final)

## [1] 556.4281

~~prelim\_final2 = glm(fracture ~ age + height + priorfrac +  
momfrac + armassist + raterisk2 +  
age\*height,  
data = glow2, family = binomial)~~

→ cat-model from slide 8

AIC(cat\_model)

## [1] 535.9395

BIC(cat\_model)

## [1] 561.2272

# Components to Assess Model Fit (I)

1. Computation and evaluation of **overall measures of fit**

2. Examination of the **individual components** of the summary statistics (often through picture)

3. Examination of other measures of the difference or **distance between the observed and fitted values**

↳ AIC / BIC & AUC - ROC  
(log/bin reg) (log-reg specific)

# Components to Assess Model Fit (I)

1. Computation and evaluation of **overall measures of fit**

LAST CLASS

2. Examination of the **individual components** of the summary statistics (often through picture)

TODAY

3. Examination of other measures of the difference or **distance between the observed and fitted values**

LAST + TODAY

# Review of Overall Measure of Fit (I)

- Test statistics of goodness-of-fit
  - When the model only contains categorical independent variables
    - Pearson residual statistic ( $X^2$ )  $\sim \chi_{J-(p+1)}^2$
    - Used when the number of covariate patterns is much less than the number of observations
  - When the model contains continuous independent variables
    - Hosmer and Lemeshow goodness-of-fit statistic ( $\hat{C}$ )  $\sim \chi_{g-2}^2$
    - Used when the number of covariate patterns is close or equal to the number of observations

# Review of Overall Measure of Fit (II)

- Area under ROC → Compare models (non-nested as well)
  - Measure the ability of the multiple logistics regression to discriminate between those subjects who experience the outcome versus those who do not
- AIC and BIC
  - Used to compare non-nested (and nested) models.

# Review of Overall Measure of Fit: GLOW Study

- Preliminary final model

$$\text{logit}(\pi(x)) = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Height}_i + \beta_3 I(\text{PF}_i = \text{"Yes"}) + \beta_4 I(\text{MF}_i = \text{"Yes"}) + \beta_5 I(\text{AA}_i = \text{"Yes"}) + \beta_6 I(\text{RR} = \text{"Greater"}) + \beta_7 \text{Age}_i \times I(\text{PF}_i = \text{"Yes"}) + \beta_8 I(\text{MF}_i = \text{"Yes"}) \times I(\text{AA}_i = \text{"Yes"})$$

PF = Prior Fracture  
MF = Mom had fracture  
AA = Arm Assistance

# Review of Overall Measure of Fit: GLOW Study

**Preliminary final model:**  $\text{logit}(\pi(x)) = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Height}_i + \beta_3 I(\text{PF}_i = \text{"Yes"}) + \beta_4 I(\text{MF}_i = \text{"Yes"}) + \beta_5 I(\text{AA}_i = \text{"Yes"}) + \beta_6 I(\text{RR} = \text{"Greater"}) + \beta_7 \text{Age}_i \times I(\text{PF}_i = \text{"Yes"}) + \beta_8 I(\text{MF}_i = \text{"Yes"}) \times I(\text{AA}_i = \text{"Yes"})$

We can examine the overall model fit using:

- Not comparing to any other models:
  - Pearson residuals: Not appropriate for this model
  - Hosmer-Lemeshow:  $\hat{C} = 6.39$ , p-value = 0.60 ✓
  - AUC-ROC: 0.7286
- Can be used to compare to other models:
  - AUC-ROC: 0.7286
  - AIC/BIC: 518.50 / 556.43

# Class 15 Learning Objectives

1. Revisit Pearson Residual: How to extract Pearson residuals and other summary measures in R
2. Apply AIC and BIC as a summary measure to make additional comparisons between potential models
3. Understand the components of calculations for logistic regression diagnostics
4. Plot and determine observations where regression does not fit well or are influential using specific diagnostic values

# Review of Number of Covariate Patterns

- Covariate patterns are the unique covariate combinations that are observed
- **For example:** model contains two categorical covariates (sex assigned at birth and smoking status), there will be **4 unique combination** of these factors  $2 \times 2 = 4$ 
  - This model has **4 *covariate patterns***
- When we have continuous covariates, the number of covariate patterns will be close to the number of individuals in the dataset

|

# Overall measure to diagnostics

- Now, we need to investigate diagnostics looking at individual data or covariate pattern data
  - Make sure the overall measure has not been influenced by certain observations
- The key quantities from logistic regression diagnostics are the components of “residual sum-of-squares”
  - The same idea as in the linear regression
  - Assessed for **each covariate pattern**  $x_j$ , by computing standardized Pearson residuals and Deviance residuals
    - Standardization using  $h_j$ , the leverage values

# Hat Matrix and Leverage Values (I)

- We have learned “hat” matrix and leverage values from linear regression diagnostics

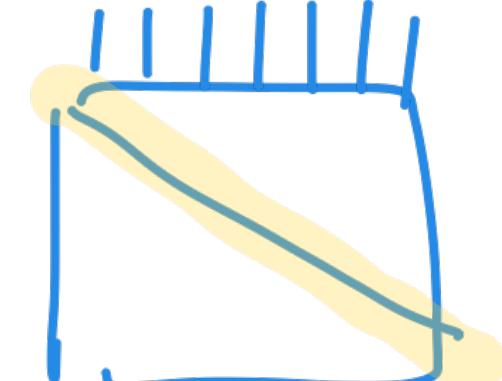
- In linear regression, the hat matrix projects the outcome variable onto the covariate space:

$$H = X(X'X)^{-1}X'; \hat{y} = H y$$

• The linear regression residuals is thus  $y - \hat{y}$ , or  $(I - H)y$

- The leverage is just the diagonal elements of the hat matrix, which is proportional to the distance of  $x_i$  to the mean of the data  $\bar{x}$

$$\hat{y}$$
$$y - H y = (I - H) y$$



# Hat Matrix and Leverage Values (II)

*j*: covariate patterns

- In logistic regression model, the hat matrix is:

$$H = V^{1/2} X (X' V X)^{-1} X' V^{1/2}$$

- The leverage is

$$h_j = m_j \times \hat{\pi}(x_j) [1 - \hat{\pi}(x_j)] x_j' (X' V X)^{-1} x_j = v_j \times b_j$$

•  $b_j$ : weighted distance of  $x_j$  from  $\bar{x}$  across covariate patterns

•  $v_j$ : model based estimator of the variance of  $y_j$

$$v_j = m_j \times \hat{\pi}(x_j) [1 - \hat{\pi}(x_j)]$$

- $h_j$  reflects the relative influence of each covariate pattern on the model's fit

# Diagnostic Statistics Computation (I)

- Two diagnostic statistics computation approach
  - Approach 1: computed by covariate pattern
    - Recommendation of HLS textbook.
    - R uses this approach
    - Identify outliers as group that shares the same covariate values (in the same covariate pattern)
  - Approach 2: individual observation approach
    - SAS uses this approach
    - Identify outliers as individual

# Diagnostic Statistics Computation (II)

- Why prefer covariate patterns approach?
  - When the number of covariate pattern is much smaller than n, there is risk that we may fail to identify influential and/or poorly fit covariate patterns using individual based on residual
- For example, consider a covariate pattern with  $m_j$  subjects, all did not have event ( $y_j = 0$ ). So the estimated logistic probability is  $\hat{\pi}_j \rightarrow \hat{\Pi}_j$
- Pearson residual computed by individual

$$r_i = -\sqrt{\frac{\hat{\pi}_j}{(1 - \hat{\pi}_j)}} \rightarrow \hat{\Pi}_j \quad \hat{\Pi}_i = \hat{\Pi}_j \quad \text{when we're within 1 covariate pattern}$$

- Pearson residual computed by covariate pattern

$$r_i = -\sqrt{m_j} \sqrt{\frac{\hat{\pi}_j}{(1 - \hat{\pi}_j)}}$$

- Difference will be large if  $m_j$  is large: usually a problem if less covariate patterns

$m_j = 1$

# 10 minute break?

# Class 15 Learning Objectives

1. Revisit Pearson Residual: How to extract Pearson residuals and other summary measures in R
2. Apply AIC and BIC as a summary measure to make additional comparisons between potential models
3. Understand the components of calculations for logistic regression diagnostics
4. Plot and determine observations where regression does not fit well or are influential using specific diagnostic values

# Diagnostics of Logistic Regression

- Model diagnostics of logistic regression can be assessed by checking how influential a covariate pattern is:
  - Look at **change in residuals if a covariate pattern is excluded**
    - Standardized Pearson residual 
    - Standardized Deviance residual 
  - Look at **change in coefficients if a covariate pattern is excluded**

# Change of Standardized Residuals

aka Pearson residual

- Change in **standardized Pearson Chi-square statistic** due to deletion of subjects with covariate pattern  $x_j$ :

$$\underline{\Delta X_j^2} = r_{sj}^2 = \frac{r_j^2}{1 - h_j}$$

high leverage<sup>then</sup>  
 $1 - h_j$  is a lower value  
means  $\Delta X_j^2(\text{std})$  increases relative to  $r_j^2$

- *Don't need to know this:* change in **standardized deviance statistic** due to deletion of subjects with covariate pattern  $x_j$ :

$$\underline{\Delta D_j} = \frac{d_j^2}{1 - h_j}$$

- Refer to Class 14 (Slide 15) for expression of  $r_j$

# Change of Estimated Coefficients

- Change in estimated coefficients due to deletion of subjects with covariate pattern  $x_j$ :

$$\underline{\Delta \hat{\beta}_j} = \frac{r_j^2 h_j}{(1 - h_j)^2}$$

- This is the logistic regression analog of Cook's influence statistic (in linear regression)

# Visual Assessment for Diagnostics of Logistic Regression (I)

- In logistic regression, we mainly rely on graphical methods
    - Because the distribution of diagnostic measures under null hypothesis (that the model fits) is only known in certain limited settings.
  - Four plots for analysis of diagnostics in logistic regression:
    - $\Delta X_j^2$  vs.  $\hat{\pi}_j$
    - $\Delta D_j$  vs.  $\hat{\pi}_j$
    - $\Delta \beta_j$  vs.  $\hat{\pi}_j$
    - $h_j$  vs.  $\hat{\pi}_j$
- 

# How do we get these values in R?

\*each row is a covariate pattern

```
diagnost_pf = dx(prelim_final)
head(diagnost_pf)
```

	(Intercept)	age	height	priorfracYes	momfracYes	armassistYes	raterisk2Greater
## 1:	1	57	171	0	0	0	0
## 2:	1	59	176	0	0	0	0
## 3:	1	57	199	0	0	0	1
## 4:	1	56	168	0	0	0	0
## 5:	1	59	170	0	0	0	0
## 6:	1	58	168	0	0	0	0
##	age:priorfracYes	momfracYes:armassistYes	y	0	0	0.04706516	0.04706516
## 1:	0	0	0	0	0	0.04200413	0.04200413
## 2:	0	0	0	0	0	0.02088401	0.02088401
## 3:	0	0	0	0	0	0.05092595	0.05092595
## 4:	0	0	0	0	0	0.05485440	0.05485440
## 5:	0	0	0	0	0	0.05675974	0.05675974
## 6:	0	0	0	0	0	0.05675974	0.05675974
##	pr	dr	h	sPr	std	sdr	dChisq
## 1:	-0.2222379	-0.3105117	0.004847631	-0.2227786	-0.3112671	0.04963029	
## 2:	-0.2093940	-0.2929567	0.005641605	-0.2099872	-0.2937866	0.04409460	
## 3:	-0.1460461	-0.2054515	0.011690129	-0.1469073	-0.2066630	0.02158174	
## 4:	-0.2316432	-0.3233217	0.004718747	-0.2321917	-0.3240872	0.05391297	
## 5:	-0.2409109	-0.3359056	0.004748526	-0.2414849	-0.3367060	0.05831496	
## 6:	-0.2453065	-0.3418603	0.004595772	-0.2458721	-0.3426486	0.06045310	
##	dDev	dBhat					
## 1:	0.09688718	0.0002417613					

observed outcome

observed covariates

$$\text{estimated prob} \rightarrow \text{# of ind in covariate pattern}$$

$$m_j \pi(\bar{x}_j)$$

n in dataframe

$$\frac{r_j^2}{(1-h_j)}$$

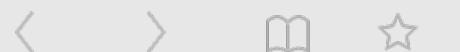
dim(diagnost\_pf)

## [1] 457 21

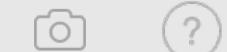
# Poll Everywhere

## Question 2

There were 500 subjects in the GLOW study dataset. Why are there 457 rows (observations) within the diagnostic summary data frame ("diagnost\_pf")?

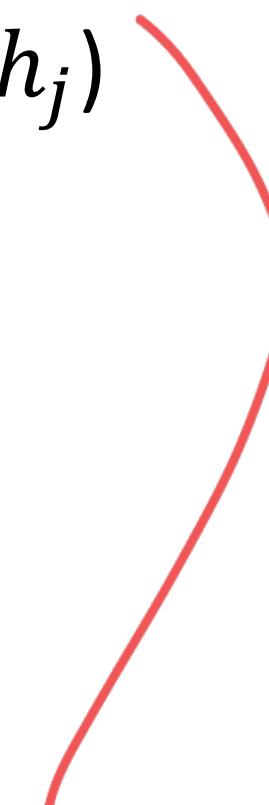


[https://www.polleverywhere.com/multiple\\_choice\\_polls/KPtNr97DDpdV82Ri7sB5B](https://www.polleverywhere.com/multiple_choice_polls/KPtNr97DDpdV82Ri7sB5B)



# Visual Assessment for Diagnostics of Logistic Regression (II)

- The plots allow us to identify those covariate patterns that are...
  - **Poorly fit**
    - Large values of  $\Delta X_j^2$  (and/or  $\Delta D_j$  if we looked at those)
  - **Influential on estimated coefficients**
    - Large values of  $\Delta \beta_j$
- If you are interested to look at the contribution of leverage ( $h_j$ ) to the values of the diagnostic statistic, you may also look at plots of:
  - $\Delta X_j^2$  vs.  $h_j$
  - $\Delta D_j$  vs.  $h_j$
  - $\Delta \beta_j$  vs.  $h_j$



# GLOW Study: Plot of $\Delta X_j^2$ vs. $\hat{\pi}_j$ (I)

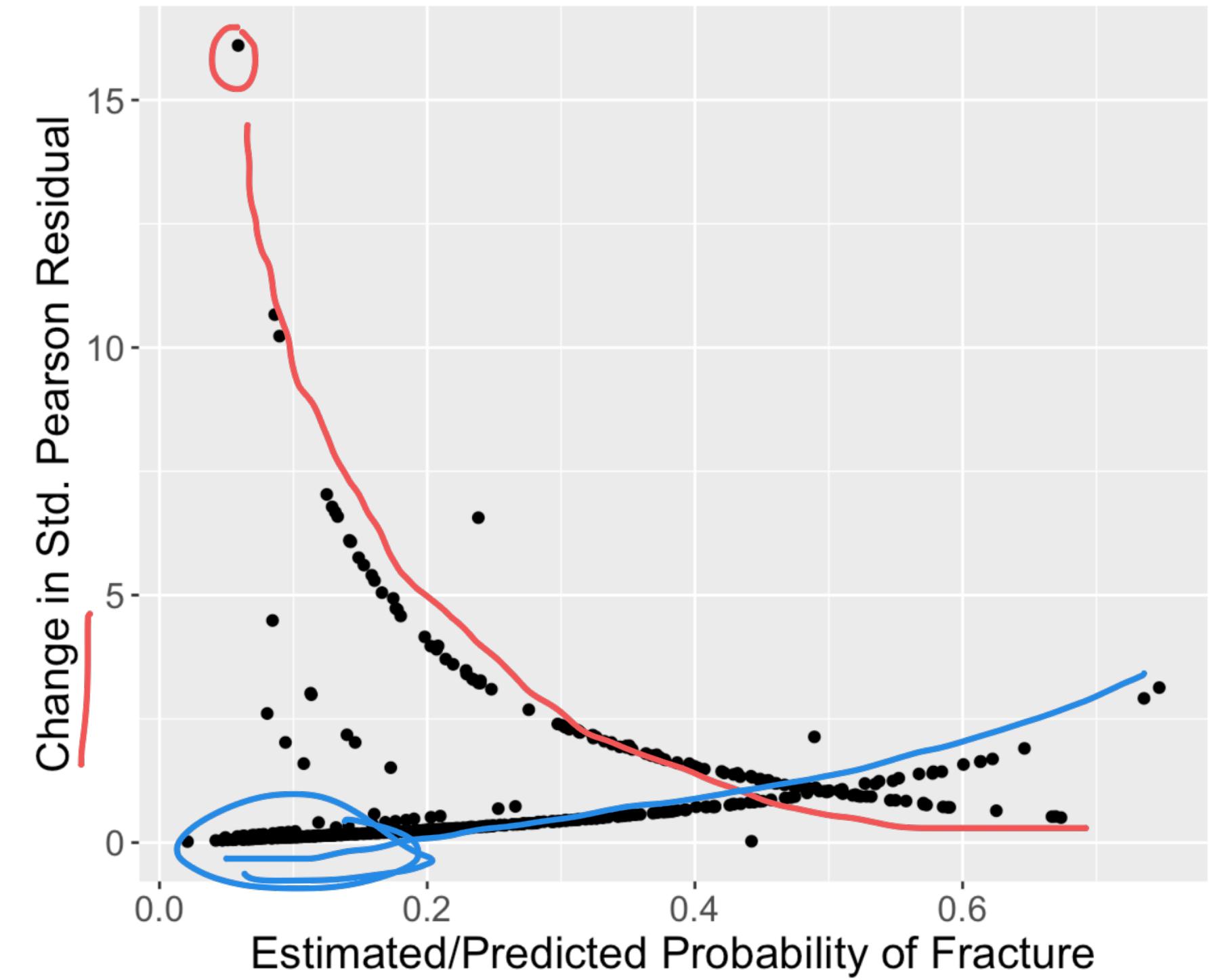
- Standardized Pearson residuals
- Generally, the points that curve from top left to bottom right of plot correspond to covariate patterns with  $y_j = 1$

- Opposite corresponds to

$y_j = 0$

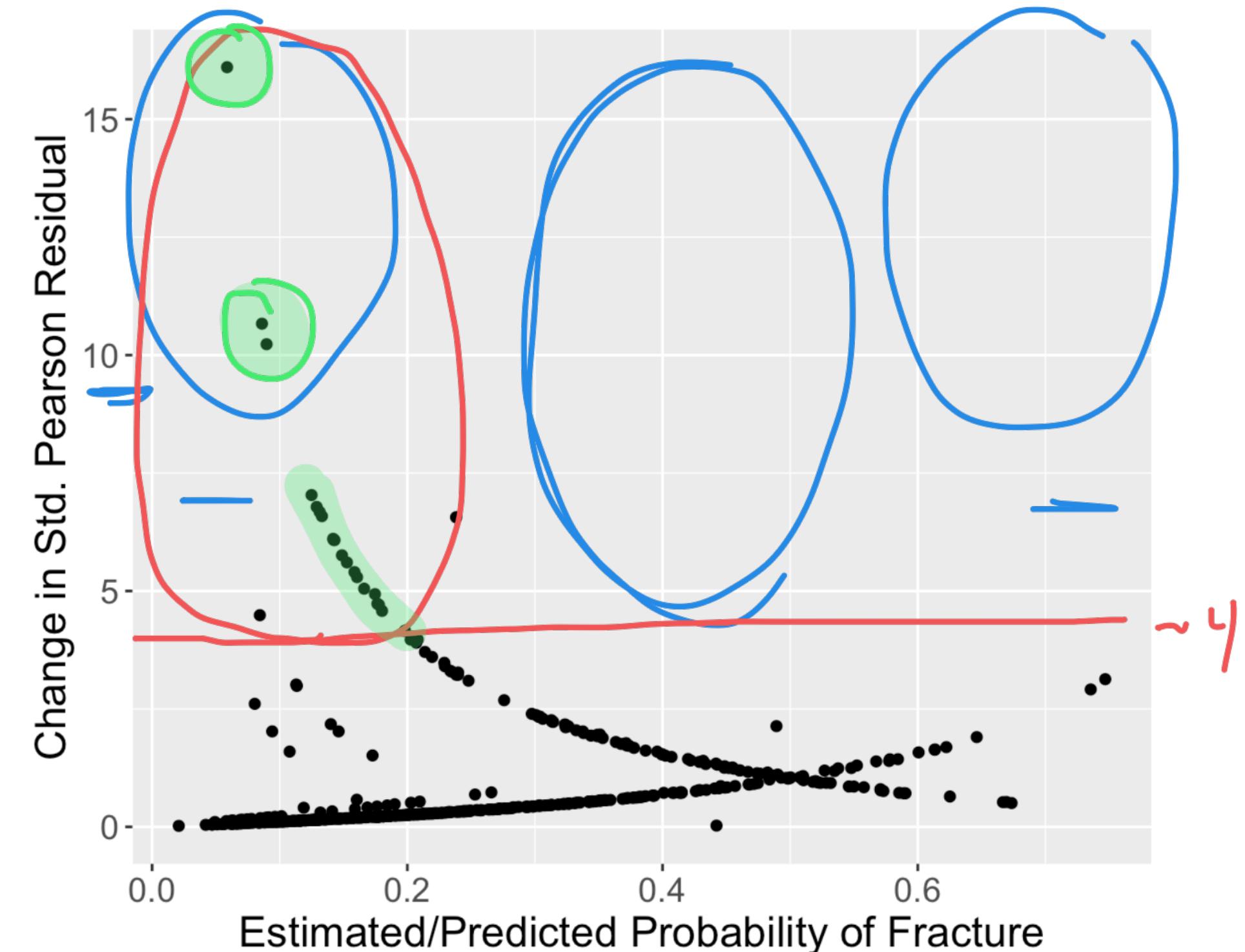
```
ggplot(diagnost_pf) + geom_point(aes(x=P, y=(sPr)^2)) +
  xlab("Estimated/Predicted Probability of Fracture") +
  ylab("Change in Std. Pearson Residual") +
  theme(text = element_text(size = 16))
```

dChisq



# GLOW Study: Plot of $\Delta X_j^2$ vs. $\hat{\pi}_j$ (II)

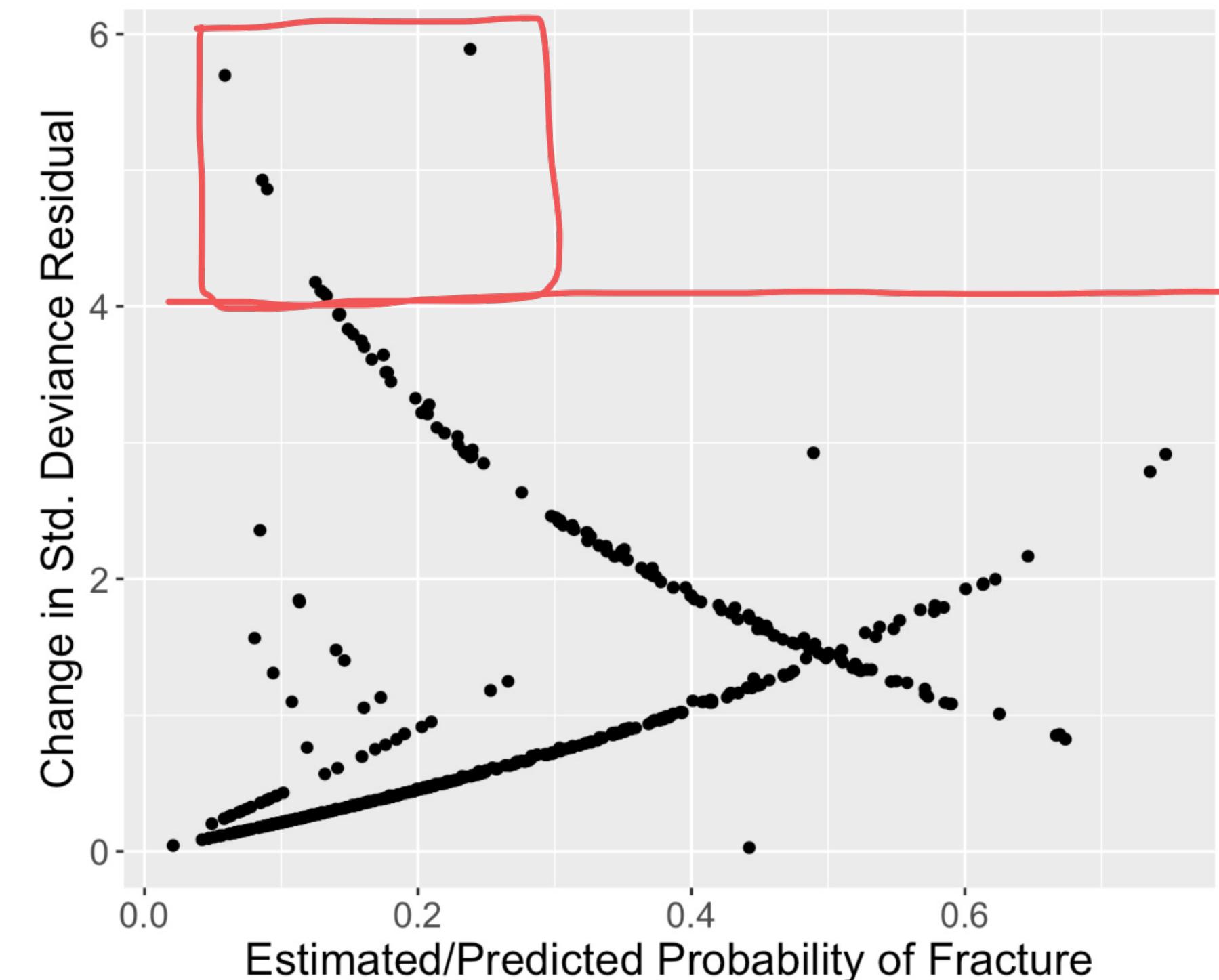
- Standardized Pearson residuals
- Points in the top left or top right corners identify the covariate patterns that are poorly fit
- We may use 4 as a crude approximation to the upper 95<sup>th</sup> percentile for  $\Delta X^2$ 
  - 95<sup>th</sup> percentile of chi-squared distribution is 3.84



```
ggplot(diagnost_pf) + geom_point(aes(x=P, y=(sPr)^2)) +  
  xlab("Estimated/Predicted Probability of Fracture") +  
  ylab("Change in Std. Pearson Residual") +  
  theme(text = element_text(size = 16))
```

# GLOW Study: Plot of $\Delta D_j$ vs. $\hat{\pi}_j$

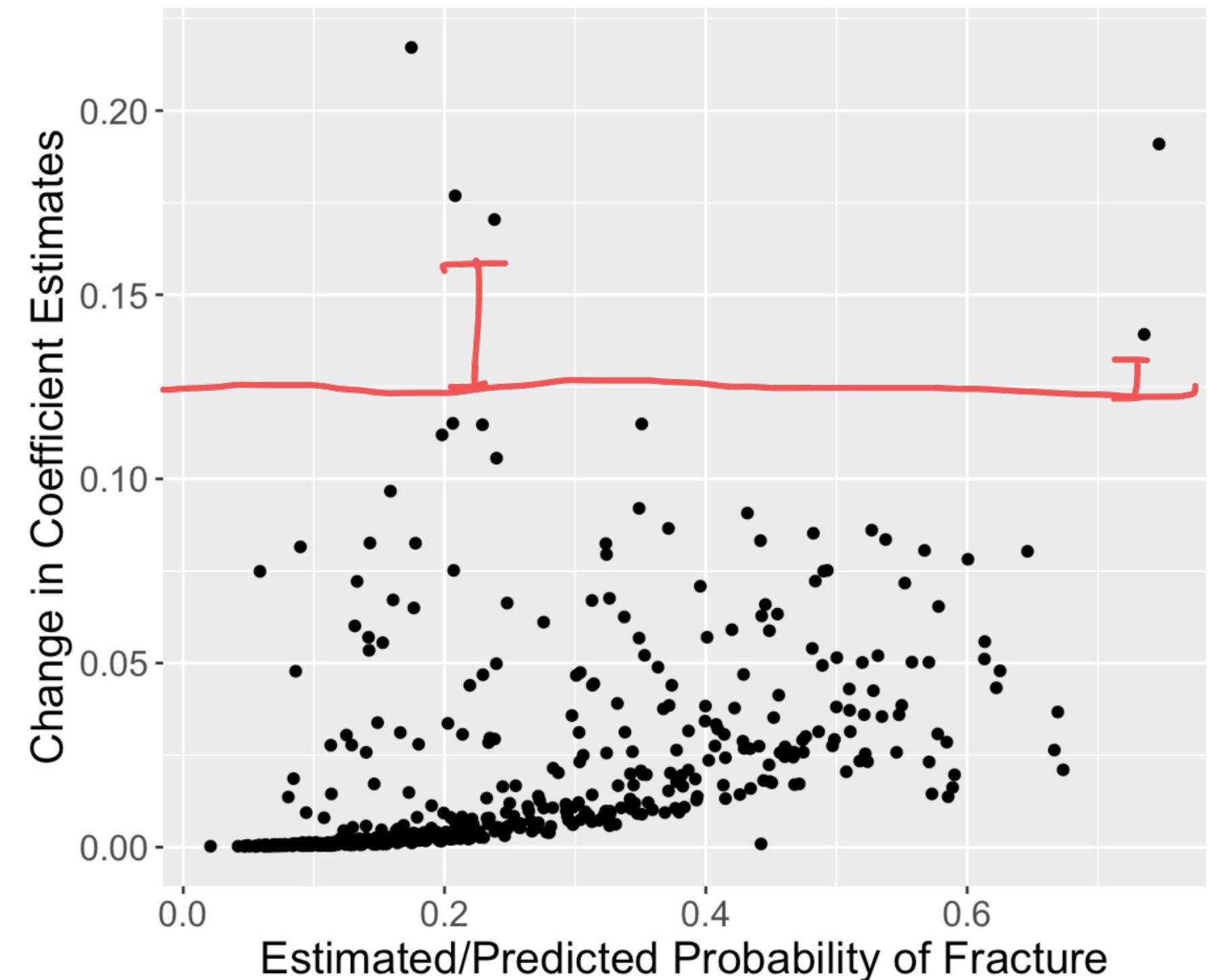
- Standardized Deviance residuals
- Same investigation as Pearson residuals
- Points in the top left or top right corners identify the covariate patterns that are poorly fit
- Use 4 as a crude approximation to the upper 95<sup>th</sup> percentile



```
ggplot(diagnost_pf) + geom_point(aes(x=P, y=(sdr)^2)) +  
  xlab("Estimated/Predicted Probability of Fracture") +  
  ylab("Change in Std. Deviance Residual") +  
  theme(text = element_text(size = 16))
```

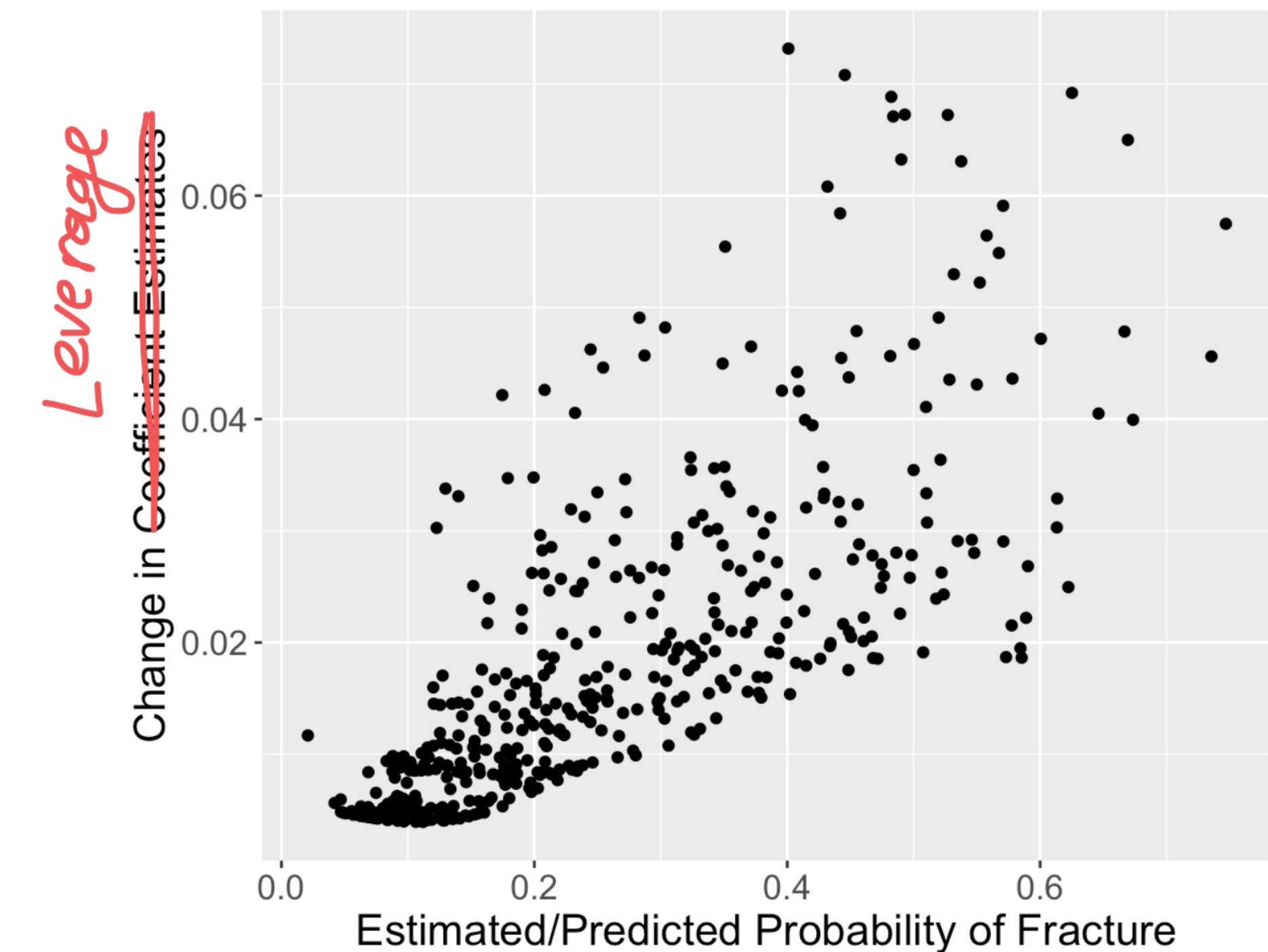
# GLOW Study: Plot of $\Delta\beta_j$ vs. $\hat{\pi}_j$

- Change in coefficient estimates
- About 5 points that are above 0.125 change in coefficients
- Book recommends flagging certain covariate patterns if change in coefficient estimates are **greater than 1**



```
ggplot(diagnost_pf) + geom_point(aes(x=P, y=dBhat)) +  
  xlab("Estimated/Predicted Probability of Fracture") +  
  ylab("Change in Coefficient Estimates") +  
  theme(text = element_text(size = 16))
```

# GLOW Study: Plot of $h_j$ vs. $\hat{\pi}_j$



# Find Out the “Influential” Observation From the Data Set

- We visually identified covariate patterns that may be poorly fit or influential
- How do we know which covariate patterns were the ones in the plot?

```
diagnost_points = diagnost_pf %>%
  mutate(Cov_patt = 1:nrow(.)) %>%
  filter((sPr)^2 > 12.5 | (sdr)^2 > 4 | dBhat > 0.125) %>%
  dplyr::select(Cov_patt, y, P, h, dChisq, dDev, dBhat) %>%
  round(., 2)
kable(diagnost_points) %>% kable_styling(bootstrap_options =
  c("striped", "hover"), full_width = F)
```

Cov_patt	y	P	h	dChisq	dDev	dBhat
340	1	0.13	0.00	6.78	4.11	0.03
350	1	0.12	0.00	7.03	4.18	0.03
391	1	0.09	0.00	10.67	4.93	0.05
412	1	0.13	0.01	6.67	4.10	0.06
426	1	0.13	0.01	6.59	4.08	0.07
428	1	0.06	0.00	16.10	5.70	0.07
436	1	0.09	0.01	10.23	4.86	0.08
453	0	0.74	0.05	2.91	2.79	0.14
454	2	0.24	0.03	6.56	5.89	0.17
455	1	0.21	0.04	3.97	3.28	0.18
456	0	0.75	0.06	3.13	2.92	0.19
457	1	0.17	0.04	4.93	3.64	0.22

# After identifying points to take out

- Take them out and rerun your logistic regression model!
- If we have time, we can switch over to R right now, and try this out
- (Did not have time to prepare a well organized slide on this)

# Three Potentials when Model Fits Poorly

- If only a few covariate pattern does not fit well ( $y_j$  differs from  $m_j \hat{\pi}_j$ ), we are not too worried
- If quite a few covariate patterns do not fit well, potential reasons can be considered:
  - The logistic model does not provide a good approximation to the correct relationship between the conditional mean  $E(Y|x_j)$  and  $x_j$ 
    - This is usually unlikely, since logistic regression model is very flexible
  - One or more important covariates missing in the model
  - At least one of the covariates in the model has been entered in the wrong scale



# Wrap-up

- 4-minute exit ticket
- Next class
  - Poisson Regression

## Class 15 Exit Ticket



<https://forms.office.com/r/SnDwJxNFZa>