# Outline

Test of Coefficients

Intervals for Coefficients

Intervals for $\pi$ at specific $X$

# Hypothesis Test for $\beta_1$

In linear regression, we computed the **test statistic**:

$$t_{obs} = \frac{\hat{\beta}_1 - 0}{\hat{se}(\hat{\beta}_1)}$$

(number of standard errors $\hat{\beta}_1$ is from 0).

$P$-value: **prob. of getting a test stat this big by chance if $H_0$ true** (i.e., $\beta_1 = 0$)

# Hypothesis Test for $\beta_1$

In logistic regression we can do the same thing, but with Normal instead of $t$ distribution.

$$z_{obs} = \frac{\hat{\beta}_1 - 0}{\hat{se}(\hat{\beta}_1)}$$

and get $P$-value: **prob of a test stat this big if $H_0$ true**

# In R

```
library(Stat2Data); data(MedGPA)
mcatModel <- glm(Acceptance ~ MCAT, data = MedGPA, family = "binomial")
summary(mcatModel) %>% coef() %>% round(3)

            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -8.712      3.236  -2.692    0.007
MCAT           0.246      0.089   2.752    0.006
```

Only 0.6% chance we'd get $\left|\hat{\beta}_1\right| \geq 0.246$ if the association is due solely to chance sampling

# Linear vs. Logistic Regression

| **Goal** | **Linear** | **Logistic** |
|---|---|---|
| Estimate coefs | Minimize SSE | Maximize Likelihood (or Minimize Deviance) |
| Check conditions | **Linearity/Const. var.**: Residual vs. Fitted **Normality**: QQ Plots | **Logit linearity**: Binned residuals vs. fitted |
| Test coefs | Measure SEs from 0, $P$-value using $t$ | Measure SEs from 0 $P$-value using Normal |

# Outline

Test of Coefficients

Intervals for Coefficients

Intervals for $\pi$ at specific $X$

# Confidence Interval for $\beta_1$

Same principle applies for confidence interval...

$$CI(\Delta\text{logit}) : \hat{\beta}_1 \pm z^* \cdot \hat{se}(\hat{\beta}_1)$$

```
confint(mcatModel) %>% round(2)

                2.5 % 97.5 %
    (Intercept) -15.77  -3.04
    MCAT          0.09   0.44
```

- But... $\beta_1$ is the rate of change of the log odds, which is hard to understand.
- More common to report a CI for **odds ratio** ($e^{\beta_1}$).

$$CI(OR) : (e^{\beta_1^{(lwr)}}, e^{\beta_1^{(upr)}})$$

# In R...

```
confint(mcatModel) %>% round(2)
```

```
                 2.5 % 97.5 %
(Intercept) -15.77  -3.04
MCAT          0.09   0.44
```

```
confint(mcatModel) %>% exp() %>% round(2)
```

```
             2.5 % 97.5 %
(Intercept)  0.00   0.05
MCAT         1.09   1.55
```

"We are 95% confident that the **odds** (**not probability**) of admittance increases by a **factor of** (is **multiplied** by) between 1.09 and 1.55 for each additional point of MCAT score"

# Linear vs. Logistic Regression

| **Goal** | **Linear** | **Logistic** |
|---|---|---|
| Estimate coefs | Minimize SSE | Maximize Likelihood |
| Check conditions | *Linearity/Const. var.*: Residual vs. Fitted *Normality*: QQ Plots | *Logit linearity*: Binned residuals vs. fitted |
| Test coefs | Measure SEs from 0, $P$-value using $t$ | Measure SEs from 0 $P$-value using Normal |
| Intervals for Params | Slope: $\beta_1$ | Odds Ratio: $e^{\beta_1}$ |

# Outline

Test of Coefficients

Intervals for Coefficients

Intervals for $\pi$ at specific $X$

# CIs at specific values

Arguably easier to interpret, CIs for $\pi$ at a few specific $X$ values

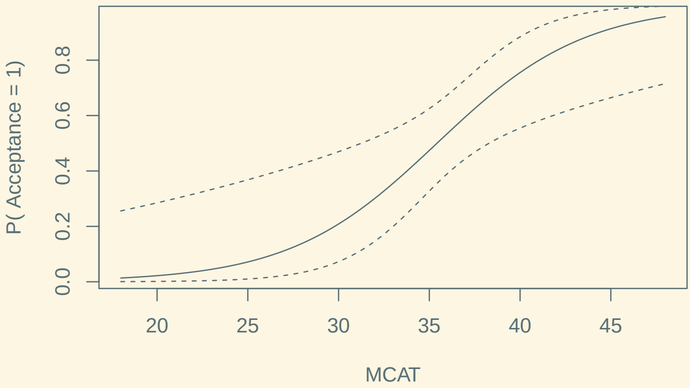
```
source("http://colindawson.net/stat213/code/helper_functions.R")
## functions made with regular makeFun() give point values but not
## intervals with logistic models, so I wrote a custom function
f_hat <- makeFun.logistic(mcatModel)
quartiles <- quantile(~MCAT, data = MedGPA)
f_hat(MCAT = quartiles, interval = "confidence", level = 0.95) %>% round(2)
```

```
     MCAT pi.hat  lwr  upr
0%     18   0.01 0.00 0.26
25%    34   0.41 0.26 0.58
50%    36   0.54 0.39 0.67
75%    39   0.71 0.52 0.84
100%   48   0.96 0.72 0.99
```

**Interpretation:** "We are 95% confident that the **probability** of acceptance for students with an MCAT score of 39 is between 52% and 84%"

# Confidence Bands

```
## Also requires sourcing helper_functions.R
## Can supply level=, xlim=, xlab= and ylab= to customize graph
plot.logistic.bands(mcatModel)
```



# Linear vs. Logistic Regression

| **Goal** | **Linear** | **Logistic** |
|---|---|---|
| Estimate coefs | Minimize SSE | Maximize Likelihood |
| Check conditions | *Linearity/Const. var.*: Residual vs. Fitted *Normality*: QQ Plots | *Logit linearity*: Binned residuals vs. fitted |
| Test coefs | Measure SEs from 0, $P$-value using $t$ | Measure SEs from 0 $P$-value using Normal |
| Intervals for Params | Slope: $\beta_1$ | Odds Ratio: $e^{\beta_1}$ |
| Intervals for Fitted Vals. | Confidence and prediction intervals | Confidence intervals only |