STA302 METHODS OF DATA ANALYSIS 1

MODULE 5: INFERENCE ON LINEAR REGRESSION COMPONENTS

PROF. KATHERINE DAIGNAULT

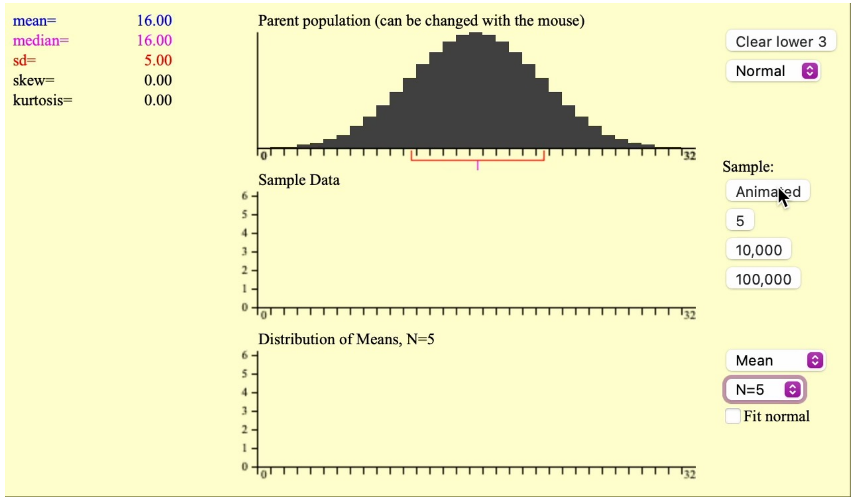
MODULE 4 OUTLINE

1. Sampling Distribution of Coefficients
2. Confidence Intervals and Hypothesis Tests for Coefficients
3. Confidence Interval for a Mean Response
4. Prediction Interval for Actual Responses

SAMPLING DISTRIBUTION REVIEW

- With any estimate, we need a measure of variation or error as it is based on data
  - Need to describe how the value varies from one sample to another
- Any time we define a sampling distribution, we utilize assumptions about the population to determine its properties
  - If assumptions don't hold, neither do these properties
- Sampling distribution is our reference for what is considered normal variation to expect
  - Pivotal quantity: $\frac{estimator - truth}{standard\ error}$ is basis of CI and hypothesis tests
  - Compared to sampling distribution to determine if estimated value reasonable or not

https://onlinestatbook.com/stat_sim/sampling_dist/



PROPERTIES OF SAMPLING DISTRIBUTIONS OF $\hat{\boldsymbol{\beta}}$

- Our assumptions say $\boldsymbol{Y}|\boldsymbol{X} \sim N(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ and our estimates are $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$
- Using linearity of Normal's, our sampling distribution is $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$
- Let's check that we get the same mean and covariance matrix if we derived them directly.

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}}|\boldsymbol{X}) &= E[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \mid \boldsymbol{X}] \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T E[\boldsymbol{Y}|\boldsymbol{X}] \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T E[\boldsymbol{X\beta} + \boldsymbol{\varepsilon} \mid \boldsymbol{X}] \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\{\boldsymbol{X\beta} + E[\boldsymbol{\varepsilon} \mid \boldsymbol{X}]\} \\
&= \underbrace{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}}_{\text{Like } c/c \,=\, 1}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}
\end{aligned}
$$

Linearity: $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

Linearity: $E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{0}$

- The LS estimators of $\boldsymbol{\beta}$ are unbiased.
- For the covariance matrix:

$$
\begin{aligned}
Cov(\hat{\boldsymbol{\beta}}|\boldsymbol{X}) &= Cov((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}|\boldsymbol{X}) \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Cov(\boldsymbol{Y}|\boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Cov(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}|\boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Cov(\boldsymbol{\varepsilon}|\boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \sigma^2\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\
&= \sigma^2\underbrace{(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}}_{\text{Like } c/c \,=\, 1}(\boldsymbol{X}^T\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}
\end{aligned}
$$

Linearity: $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

Constant variance & uncorrelated errors

Theorem 3.6d. Let $\mathbf{z} = \mathbf{Ay}$ and $\mathbf{w} = \mathbf{By}$, where $\mathbf{A}$ is a $k \times p$ matrix of constants, $\mathbf{B}$ is an $m \times p$ matrix of constants, and $\mathbf{y}$ is a $p \times 1$ random vector with covariance matrix $\boldsymbol{\Sigma}$. Then

(i) $\operatorname{cov}(\mathbf{z}) = \operatorname{cov}(\mathbf{Ay}) = \mathbf{A\Sigma A}'$, (3.44)

MORE ABOUT COVARIANCE MATRICES

- Covariance matrices combine information about
  - Variance of each individual random variable
  - How any two random variables vary together
  - E.g. $Cov(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$
- We read the matrix information as

$$Cov(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \begin{pmatrix} Var(\varepsilon_1|\boldsymbol{X}) & Cov(\varepsilon_1, \varepsilon_2|\boldsymbol{X}) & \cdots & Cov(\varepsilon_1, \varepsilon_n|\boldsymbol{X}) \\ Cov(\varepsilon_1, \varepsilon_2|\boldsymbol{X}) & Var(\varepsilon_2|\boldsymbol{X}) & \cdots & Cov(\varepsilon_2, \varepsilon_n|\boldsymbol{X}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\varepsilon_1, \varepsilon_n|\boldsymbol{X}) & Cov(\varepsilon_2, \varepsilon_n|\boldsymbol{X}) & \cdots & Var(\varepsilon_n|\boldsymbol{X}) \end{pmatrix}$$

- This is still the case with the covariance matrix in the sampling distribution:

$$Cov(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \begin{pmatrix} Var(\hat{\beta}_0|\boldsymbol{X}) & Cov(\hat{\beta}_0, \hat{\beta}_1|\boldsymbol{X}) & \cdots & Cov(\hat{\beta}_0, \hat{\beta}_p|\boldsymbol{X}) \\ Cov(\hat{\beta}_0, \hat{\beta}_1|\boldsymbol{X}) & Var(\hat{\beta}_1|\boldsymbol{X}) & \cdots & Cov(\hat{\beta}_1, \hat{\beta}_p|\boldsymbol{X}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_0, \hat{\beta}_p|\boldsymbol{X}) & Cov(\hat{\beta}_1, \hat{\beta}_p|\boldsymbol{X}) & \cdots & Var(\hat{\beta}_p|\boldsymbol{X}) \end{pmatrix}$$

- MLR elements are not easily expressed
- SLR elements are much easier to understand:

$$Cov(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}\begin{pmatrix} \sum x_i^2/n & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

- The estimated coefficients have non-zero covariance, showing everything is conditional in regression.

UNKNOWN ERROR VARIANCE

- The sampling distribution has one hiccup.
- $\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$ contains an unknown parameter σ^2
  - Inference requires knowledge of the variation in the possible estimates
  - Knowing the form of the variance matrix $\sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ is not enough
  - Need a value to compute margins of error or standardized test statistics
- Need an estimate of σ^2 to use the sampling distribution in practice
- Estimate σ^2 with $s^2 = \frac{RSS}{n-p-1} = \frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}}{n-p-1}$
  - E.g. in simple regression $\hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}} = \sum(y_i \ - \hat{\beta}_0 \ - \hat{\beta}_1 \ x_i \ )^2$
  - Like sample variance, denominator is degrees of freedom (i.e. observations (n) minus how many parameters were estimated in the numerator ($p+1$))
- In practice, we now use $s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ as variance matrix
  - But this adds additional sampling variation
- A related distribution is used to adjust for this:

$$\frac{\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}}{\sqrt{s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}}} \sim T_{n-p-1}$$

MODULE 4 OUTLINE

1. Sampling Distribution of Coefficients
2. Confidence Intervals and Hypothesis Tests for Coefficients
3. Confidence Interval for a Mean Response
4. Prediction Interval for Actual Responses

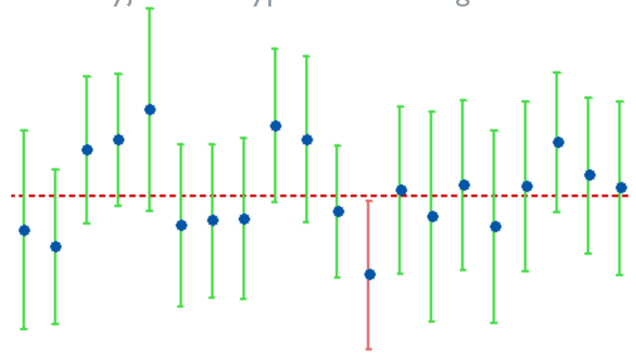
GENERAL CI AND TEST FORM AND USAGE

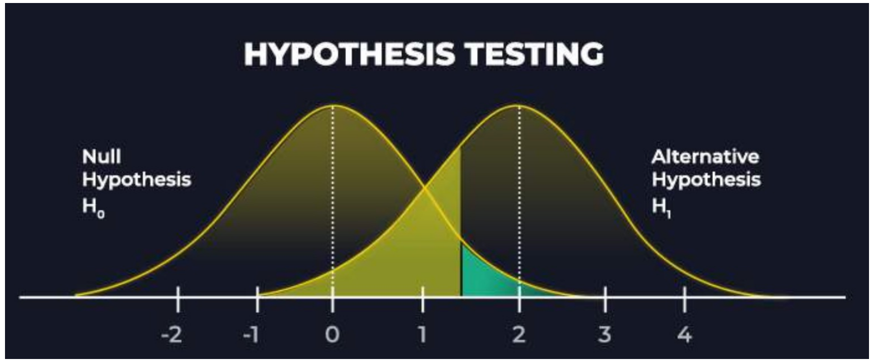
- Many inferential processes take a common form:
  - CIs: estimate – (critical value)*(standard error)
  - Test statistics: $\frac{\text{estimate} - \text{truth}}{\text{standard error}}$
- Sampling distribution has all these components:

$$\frac{\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}}{\sqrt{s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}}} \sim T_{n-p-1}$$

- CIs: chance your sample was one of the $(1-\alpha)\%$ CIs that overlapped the truth
- Tests: chance that you would have estimated an even more extreme value farther from the truth.

https://statisticsbyjim.com/hypothesis-testing/confidence-interval/





https://www.analyticssteps.com/blogs/what-hypothesis-testing-types-and-methods

INFERENCE ON INDIVIDUAL COEFFICIENTS β_j

$(1-\alpha)\%$ confidence interval for β_j:

$$\hat{\beta}_j \pm t_{\frac{\alpha}{2},n-p-1}\, s\sqrt{(\boldsymbol{X}^T\boldsymbol{X})^{-1}_{(j+1,j+1)}}$$

- α is our chosen significance level, while $1-\alpha$ is our confidence level
- The critical value corresponds to the $^{\alpha}/_{2}$ quantile of the T distribution with $n-p-1$ degrees of freedom
- $(\boldsymbol{X}^T\boldsymbol{X})^{-1}{}_{(j+1,j+1)}$ refers to $j+1$ element of the main diagonal
  - E.g., $Var(\hat{\beta}_0) = s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}{}_{(1,1)}$ whereas $Var(\hat{\beta}_3) = s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}{}_{(4,4)}$

Hypothesis Test of H_0: $\beta_j = \beta_j^0$: $t^* = \dfrac{\hat{\beta}_j - \beta_j^0}{s\sqrt{(\boldsymbol{X}^T\boldsymbol{X})^{-1}_{(j+1,j+1)}}}$

- β_j^0 is the value we hypothesize is the truth (usually 0)
- Standard error $s\sqrt{(\boldsymbol{X}^T\boldsymbol{X})^{-1}_{(j+1,j+1)}}$ used to standardize the difference between estimate and hypothesized value
  - Compare this difference to expected variability to see if hypothesized value is plausible.
- Use same statistics regardless of which $\hat{\beta}_j$ or which value β_j^0 we test, or even which alternative (H_A: $\beta_j \neq \beta_j^0$ or H_A: $\beta_j > \beta_j^0$ or H_A: $\beta_j < \beta_j^0$)

CONCLUDING INFERENCE ON COEFFICIENTS

$(1-\alpha)\%$ CI for β_j: $\hat{\beta}_j \pm t_{\frac{\alpha}{2},n-p-1}\, s\sqrt{(\boldsymbol{X}^T\boldsymbol{X})^{-1}_{(j+1,j+1)}}$

- Built using data and sampling distribution
- Window that is 2 $t_{\frac{\alpha}{2},n-p-1}$ standard errors wide, centered on estimate $\hat{\beta}_j$
  - Would see the true β_j in this window $(1-\alpha)\%$ of the time
- Interpretation of interval: $(1-\alpha)\%$ of all intervals computed using data repeatedly obtained from the same population would contain the true β_j.
  - Can also say it represents plausible values of β_j with $(1-\alpha)\%$ confidence.

Hypothesis Test of H_0: $\beta_j = \beta_j^0$: $t^* = \dfrac{\hat{\beta}_j - \beta_j^0}{s\sqrt{(\boldsymbol{X}^T\boldsymbol{X})^{-1}_{(j+1,j+1)}}}$

- Testing H_0: $\beta_j = 0$ versus H_A: $\beta_j \neq 0$ is most common
  - Tests the null that no linear relationship exists between X_j and Y (in the presence of other predictors).
- Conclude the test by comparing to sampling distribution:
  - If $|t^*| > t_{\frac{\alpha}{2},n-p-1}$, then reject the null
  - If $P\left(|T_{n-p-1}| \geq |t^*|\right) < \alpha$, then reject the null
  - Claim a significant linear relationship exists

EXAMPLE BY HAND

The estimated simple linear model relating Number of Rooms Cleaned (Y) to the Size of the Cleaning Crew (X) is

$$\hat{y}_i = 1.785 + 3.701 x_i.$$

We have

- 53 observations
- a sample mean of

  $\bar{x} = 8.679$
- sample variance of

  $s_x^2 = 23.068 = \frac{\sum(x_i - \bar{x})^2}{53-1}$
- an $RSS = 2744.796$.

1. Find variances $s^2 = \frac{RSS}{n-p-1} = \frac{2744.796}{53-1-1} = 53.82$

```
> X <- as.matrix(cbind(rep(1, 53), data[,2]))
> solve(t(X) %*% X)
            [,1]          [,2]
[1,]  0.081666037 -0.0072354348
[2,] -0.007235435  0.0008336479
```

$= (\boldsymbol{X}^T\boldsymbol{X})^{-1}$

$$Var(\hat{\beta}_0) = s^2 (\boldsymbol{X}^T\boldsymbol{X})^{-1}{}_{(1,1)} = 53.82(0.8166\ldots) \cong 4.395$$

$$Var(\hat{\beta}_1) = s^2 (\boldsymbol{X}^T\boldsymbol{X})^{-1}{}_{(2,2)} = 53.82(0.00083\ldots) \cong 0.0449$$

Can also use algebraic formulae in SLR, based on slide 5:

$$Var(\hat{\beta}_0) = \frac{s^2 \sum x_i^2 / n}{\sum(x_i - \bar{x})^2} = s^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right) = 53.82\left(\frac{1}{53} + \frac{8.679^2}{(53-1)(23.068)}\right) \cong 4.395$$

$$Var(\hat{\beta}_1) = \frac{s^2}{\sum(x_i - \bar{x})^2} = \frac{53.82}{(53-1)(23.068)} \cong 0.0449$$

2. Critical value/distribution Sampling distribution is $T_{n-p-1} = T_{51}$ $\Longrightarrow$ $\left|t_{\frac{0.05}{2},51}\right| \cong 2.00$

3. Compute Interval/Test Statistic

$$\hat{\beta}_0 \pm t_{\frac{0.05}{2},51}\sqrt{s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}{}_{(1,1)}} = 1.785 \pm 2.00\sqrt{4.395} = [-2.41, 5.98]$$

One of the 95% of all intervals computed using data repeatedly obtained from the same population that contains the true β_0

$$H_0: \beta_1 = 0 \text{ vs } H_A: \beta_1 \neq 0: \frac{\hat{\beta}_1 - 0}{\sqrt{s^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}{}_{(2,2)}}} = \frac{3.701}{\sqrt{0.0449}} = 17.466$$

Since 17.466 > 2.00, reject H_0, conclude a statistically significant linear relationship exists

EXAMPLE USING R

The estimated simple linear model relating Number of Rooms Cleaned (Y) to the Size of the Cleaning Crew (X) is

$$\hat{y}_i = 1.785 + 3.701 x_i.$$

We have

- 53 observations
- a sample mean of

  $\bar{x} = 8.679$
- sample variance of

  $s_x^2 = 23.068 = \frac{\sum(x_i - \bar{x})^2}{53-1}$
- an $RSS = 2744.796.$

Fit the model and look at the summary:

```
> # fit model first
> model <- lm(Rooms ~ Crews, data=data)
> # use summary() function to view test statistics
> summary(model)

Call:
lm(formula = Rooms ~ Crews, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-15.9990  -4.9901   0.8046   4.0010  17.0010

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.7847     2.0965   0.851    0.399
Crews         3.7009     0.2118  17.472   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.336 on 51 degrees of freedom
Multiple R-squared:  0.8569,    Adjusted R-squared:  0.854
F-statistic: 305.3 on 1 and 51 DF,  p-value: < 2.2e-16
```

$\hat{\beta}_j$ $\quad \sqrt{Var(\hat{\beta}_j)}$ $\quad t^*$ $\quad P(|T_{n-p-1}| \geq |t^*|)$ $\quad \sqrt{s^2}$

Intercept not significantly different from 0 (fail to reject)

Since p-value < 0.05, reject H_0, conclude a statistically significant linear relationship exists

Confidence Intervals:

```
> # use confint() function to get confidence intervals
> confint(model)
                2.5 %   97.5 %
(Intercept) -2.424160 5.993558
Crews        3.275653 4.126134
> # can change the confidence level
> confint(model, level=0.9)
                  5 %     95 %
(Intercept) -1.727503 5.296900
Crews        3.346039 4.055748
```

Default is 95% but can change

MODULE 4 OUTLINE

1. Sampling Distribution of Coefficients
2. Confidence Intervals and Hypothesis Tests for Coefficients
3. Confidence Interval for a Mean Response
4. Prediction Interval for Actual Responses

SAMPLING DISTRIBUTION OF $\hat{E}(Y|X)$

- Can also make inference on $E(\boldsymbol{Y}|\boldsymbol{X})$, estimated by $\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$
  - $E(\boldsymbol{Y}|\boldsymbol{X})$ is a parameter so we can build CIs and tests
- Treat each mean response individually
  - For each set of values $\boldsymbol{x}_0^T = (1, x_1, x_2, \ldots, x_p)$, estimate $\hat{y}_0 = \hat{E}(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}_0{}^T) = \boldsymbol{x}_0^T\widehat{\boldsymbol{\beta}}$, a single estimated mean.
- Sampling distribution of $\hat{y}_0$ uses similar ideas:
  - $\hat{y}_0$ is a linear combination of $\boldsymbol{Y}$: $\hat{y}_0 = \boldsymbol{x}_0^T\widehat{\boldsymbol{\beta}} = \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$
  - Use linearity of Normals to get

$$\hat{y}_0|\boldsymbol{X}, \boldsymbol{x}_0 \sim N(\boldsymbol{x}_0^T\boldsymbol{\beta}, \sigma^2\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0)$$

- If assumptions hold, then $\hat{y}_0$ = is unbiased:

$$E(\hat{y}_0|\boldsymbol{X}, \boldsymbol{x}_0) = E\left(\boldsymbol{x}_0^T\widehat{\boldsymbol{\beta}}\middle|\boldsymbol{X}, \boldsymbol{x}_0\right) = \boldsymbol{x}_0^T E\left(\widehat{\boldsymbol{\beta}}\middle|\boldsymbol{X}, \boldsymbol{x}_0\right) = \boldsymbol{x}_0^T\boldsymbol{\beta}$$

- If assumptions hold, the covariance matrix is

$$\begin{aligned} Cov(\hat{y}_0|\boldsymbol{X}, \boldsymbol{x}_0) &= Cov\left(\boldsymbol{x}_0^T\widehat{\boldsymbol{\beta}}\middle|\boldsymbol{X}, \boldsymbol{x}_0\right) \\ &= \boldsymbol{x}_0^T Cov\left(\widehat{\boldsymbol{\beta}}\middle|\boldsymbol{X}, \boldsymbol{x}_0\right)\boldsymbol{x}_0 \\ &= \sigma^2\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0 \end{aligned}$$

- As before, σ^2 must be estimated by s^2 giving

$$\frac{\hat{y}_0 - \boldsymbol{x}_0^T\boldsymbol{\beta}}{\sqrt{s^2\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0}} \sim T_{n-p-1}$$

INFERENCE ON MEAN RESPONSE $\boldsymbol{x}_0^T\boldsymbol{\beta}$

$(1-\alpha)\%$ confidence interval for $y_0 = \boldsymbol{x}_0^T\boldsymbol{\beta}$:

$$\boldsymbol{x}_0^T\widehat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2},n-p-1}\, s\sqrt{\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0}$$

- Same T_{n-p-1} distribution as before
- Window that is 2 $t_{\frac{\alpha}{2},n-p-1}$ standard errors wide, centered on estimate $\boldsymbol{x}_0^T\widehat{\boldsymbol{\beta}}$
- If working with simple linear model, can use below for standard error instead:

$$\sqrt{Var(\hat{y}_0|X,x_0)} = \sqrt{s^2\left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right)}$$

Hypothesis Test of H_0: $y_0 = y_0^0$: $t^* = \dfrac{\hat{y}_0 - y_0^0}{s\sqrt{\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0}}$

- Can conduct hypothesis test on mean response
  - It's a parameter so we can always test a specific value
  - Not common
- Conclude based on sampling distribution T_{n-p-1}
  - If $|t^*| > t_{\frac{\alpha}{2},n-p-1}$, then reject the null
  - If $P\left(\left|T_{n-p-1}\right| \geq |t^*|\right) < \alpha$, then reject the null
- No default value or special interpretation

EXAMPLE BY HAND & WITH R

The estimated simple linear model relating Number of Rooms Cleaned (Y) to the Size of the Cleaning Crew (X) is

$$\hat{y}_i = 1.785 + 3.701 x_i.$$

We have

- 53 observations
- a sample mean of
  $\bar{x} = 8.679$
- sample variance of
  $s_x^2 = 23.068 = \frac{\sum(x_i - \bar{x})^2}{53-1}$
- an $RSS = 2744.796$.

Let's estimate mean response for 5 Crews: $\boldsymbol{x}_0^T = (1 \quad 5) \implies \hat{y}_0 = (1 \quad 5)\begin{pmatrix}1.785\\3.701\end{pmatrix} = 20.29$

For variance of estimate, need to use whole $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ matrix:

$$Var(\hat{y}_0) = s^2 \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0 = 53.82 \times (1 \quad 5)\begin{pmatrix}0.081666037 & -0.0072354348\\-0.0072354348 & 0.0008336479\end{pmatrix}\begin{pmatrix}1\\5\end{pmatrix} \cong 1.623$$

Alternatively, use algebraic form in SLR:

$$Var(\hat{y}_0) = s^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right) = 53.82\left(\frac{1}{53} + \frac{(5 - 8.679)^2}{(53-1)(23.068)}\right) \cong 1.623$$

Critical value/distribution Sampling distribution is $T_{n-p-1} = T_{51} \implies \left|t_{\frac{0.05}{2},51}\right| \cong 2.00$

95% Confidence interval for mean response:

$$\boldsymbol{x}_0^T\hat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2},n-p-1}\, s\sqrt{\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0} = 20.29 \pm 2.00\sqrt{1.623} = [17.74, 22.84]$$

```
> # create new data to predict at
> new <- data.frame(Crews=5)
>
> # use predict() with model from before
> predict(model, newdata=new, interval="confidence", level=0.95)
       fit      lwr      upr
1 20.28917 17.73171 22.84662
```

What value to predict at

default

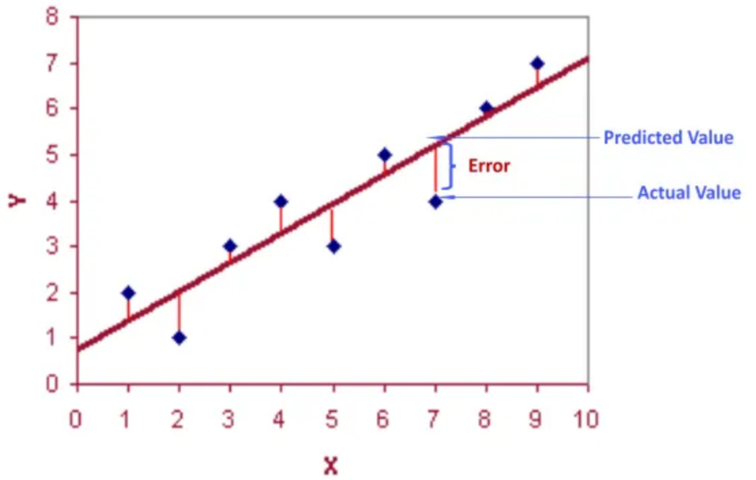
MODULE 4 OUTLINE

1. Sampling Distribution of Coefficients
2. Confidence Intervals and Hypothesis Tests for Coefficients
3. Confidence Interval for a Mean Response
4. Prediction Interval for Actual Responses

DIFFERENCE BETWEEN ACTUAL Y AND MEAN RESPONSE

- Regression model gives predicted values ONLY for $E(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}_0)$
  - We only ever get values that are estimates of conditional mean responses.
  - $E(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}_0)$ is a parameter, a fixed but unknown value
- Want to make a prediction about an individual person and their respective y_0
  - Actual individual's response is a realization of a random variable Y in the population
  - It can take any number of possible values when $\boldsymbol{X} = \boldsymbol{x}_0$
  - It also likely is not equivalent to $E(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}_0)$
    - i.e. does not necessarily lie on the regression surface

https://medium.com/@mygreatlearning/rmse-what-does-it-mean-2d446c0b1d0e



- Need to account for difference between actual value (what we want) and predicted value (what regression gives) using prediction error

$$y_0 - \hat{y}_0 = \boldsymbol{x}_0^T\boldsymbol{\beta} + \varepsilon_0 - \hat{y}_0 = (\boldsymbol{x}_0^T\boldsymbol{\beta} - \hat{y}_0) + \varepsilon_0$$

DISTRIBUTION OF PREDICTION ERROR

- We use $\hat{y}_0 = \boldsymbol{x}_0^T\widehat{\boldsymbol{\beta}}$ as our predicted value
- $y_0 - \hat{y}_0 = \boldsymbol{x}_0^T\boldsymbol{\beta} + \varepsilon_0 - \hat{y}_0 = (\boldsymbol{x}_0^T\boldsymbol{\beta} - \hat{y}_0) + \varepsilon_0$
  - Says the error in our prediction depends on the error in the population (ε_0) and how well we estimate our mean response ($\boldsymbol{x}_0^T\boldsymbol{\beta} - \hat{y}_0$)
- Our prediction error on average is 0:

$$E(y_0 - \hat{y}_0|X, x_0) = \boldsymbol{x}_0^T\boldsymbol{\beta} - E(\hat{y}_0|X, x_0) = \boldsymbol{x}_0^T\boldsymbol{\beta} - \boldsymbol{x}_0^T\boldsymbol{\beta}$$

  - Using properties of $\hat{y}_0$ we worked with earlier
- When assumptions hold, y_0 and $\hat{y}_0$ are uncorrelated and should be sampled randomly from population

- The variance of the prediction error is

$$\begin{aligned} Var(y_0 - \hat{y}_0|\boldsymbol{X}, \boldsymbol{x}_0) &= Var(\boldsymbol{x}_0^T\boldsymbol{\beta} + \varepsilon_0 - \boldsymbol{x}_0^T\widehat{\boldsymbol{\beta}}|\boldsymbol{X}, \boldsymbol{x}_0) \\ &= Var(\varepsilon_0|\boldsymbol{X}, \boldsymbol{x}_0) + Var(\boldsymbol{x}_0^T\widehat{\boldsymbol{\beta}}|\boldsymbol{X}, \boldsymbol{x}_0) \\ &= \sigma^2 + \sigma^2\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0 \\ &= \sigma^2[1 + \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0] \end{aligned}$$

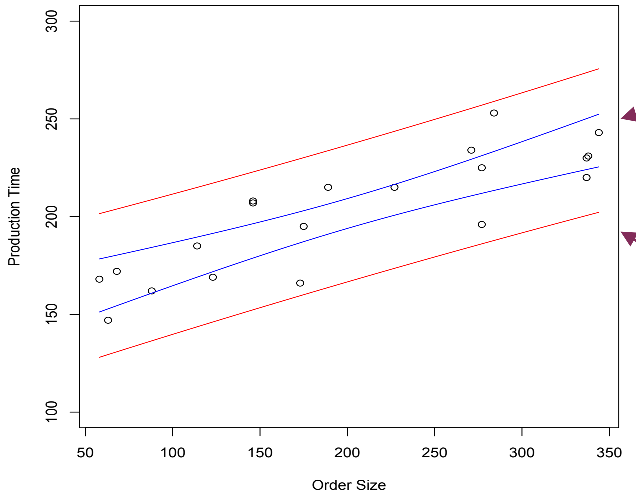
- As both $\hat{y}_0$ and y_0 are responses and should be Normal, we get the distribution of prediction error

$$y_0 - \hat{y}_0|\boldsymbol{X}, \boldsymbol{x}_0 \sim N(0, \sigma^2[1 + \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0])$$

- Since σ^2 is unknown, we estimate with s^2
  - Means T_{n-p-1} distribution describes prediction error better than Normal.

PREDICTION INTERVAL FOR ACTUAL RESPONSE

- Since we can't predict an actual value, we instead create a prediction interval
  - Gives a range of possible values for an actual response
  - Not the same as a confidence interval because we don't estimate a parameter
- The interval is centered at our estimated mean response $\hat{y}_0$
- It is also wider than the CI for $E(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{x}_0)$ due to extra σ^2
  - Includes variation in estimating conditional mean, and variation around conditional mean.
  - most likely $(1-\alpha)\%$ of response values random variable could take



$(1-\alpha)\%$ PI: $\boldsymbol{x}_0^T\hat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2},n-p-1}\, s\sqrt{1+\boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0}$

For SLR: $\hat{\beta}_0+\hat{\beta}_1x_0 \pm t_{\frac{\alpha}{2},n-2}\, s\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$

EXAMPLE BY HAND & USING R

The estimated simple linear model relating Number of Rooms Cleaned (Y) to the Size of the Cleaning Crew (X) is

$$\hat{y}_i = 1.785 + 3.701x_i.$$

We have

- 53 observations
- a sample mean of

  $\bar{x} = 8.679$
- sample variance of

  $s_x^2 = 23.068 = \frac{\sum(x_i-\bar{x})^2}{53-1}$
- an $RSS = 2744.796.$

Predict an actual response for 5 Crews: $\boldsymbol{x}_0^T = (1 \quad 5) \implies \hat{y}_0 = (1 \quad 5)\begin{pmatrix}1.785\\3.701\end{pmatrix} = 20.29$

For variance, can do matrix multiplication again, or realize we add one more s^2:

$$Var(\hat{y}_0) = s^2[1 + \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0] = 53.82 + 1.623 = 55.44$$

Same for algebraic form in SLR:

$$Var(\hat{y}_0) = s^2\left(1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right) = 53.82\left(1 + \frac{1}{53} + \frac{(5-8.679)^2}{(53-1)(23.068)}\right) \cong 55.44$$

Critical value/distribution Sampling distribution is $T_{n-p-1} = T_{51} \implies \left|t_{\frac{0.05}{2},51}\right| \cong 2.00$

95% Prediction interval for actual response:

$$\boldsymbol{x}_0^T\widehat{\boldsymbol{\beta}} \pm t_{\frac{\alpha}{2},n-p-1}\, s\sqrt{\mathbf{1} + \boldsymbol{x}_0^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_0} = 20.29 \pm 2.00\sqrt{55.44} \cong [5.40, 35.18]$$

```
>
> # create new data to predict at
> new <- data.frame(Crews=5)
> # use predict() with model from before
> predict(model, newdata=new, interval="prediction", level=0.95)
       fit      lwr      upr
1 20.28917 5.340774 35.23756
```

Only change is to say you want "prediction" interval, not "confidence" interval

MODULE TAKE-AWAYS

1. How did we determine the properties of the sampling distribution and where did assumptions play a role?
2. Why do we use a T distribution when working with the sampling distribution in practice?
3. How do we compute confidence/prediction intervals and conduct hypothesis tests on regression components?
4. How are the inferential procedures concluded?
5. What is the difference between estimating a mean response and predicting an actual response?