Linear and Generalized Linear Models (4433LGLM6Y)

Statistical Inference

Meeting 5

Vahe Avagyan

Biometris, Wageningen University and Research

Statistical inference

- inference for individual coefficients: t-tests and confidence intervals
- inference for several coefficients: F-tests
- general linear hypotheses

Linear Model Theory

- Linear model: Reminder

$$\boldsymbol{y} = \mathbf{X\beta} + \epsilon,$$

where $\epsilon \sim N_n(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_n)$ and $\mathbf{X}_{n\times(k+1)}$ is the model matrix.

- Fitting the model to data gives the vectors of fitted values and residuals:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e},$$

- Normal equations to obtain the LS estimators $\mathbf{b}$ of $\boldsymbol{\beta}$:

$$(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{y}.$$

Distribution of least-squares estimator

- LS estimator:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

- Recall the properties .

1. $\mathbf{b}$ is a linear estimator: $\mathbf{b} = \mathbf{M}\mathbf{y}$, for some $\mathbf{M}$

2. $\mathbf{b}$ is an unbiased estimator: $E(\mathbf{b}) = \beta$

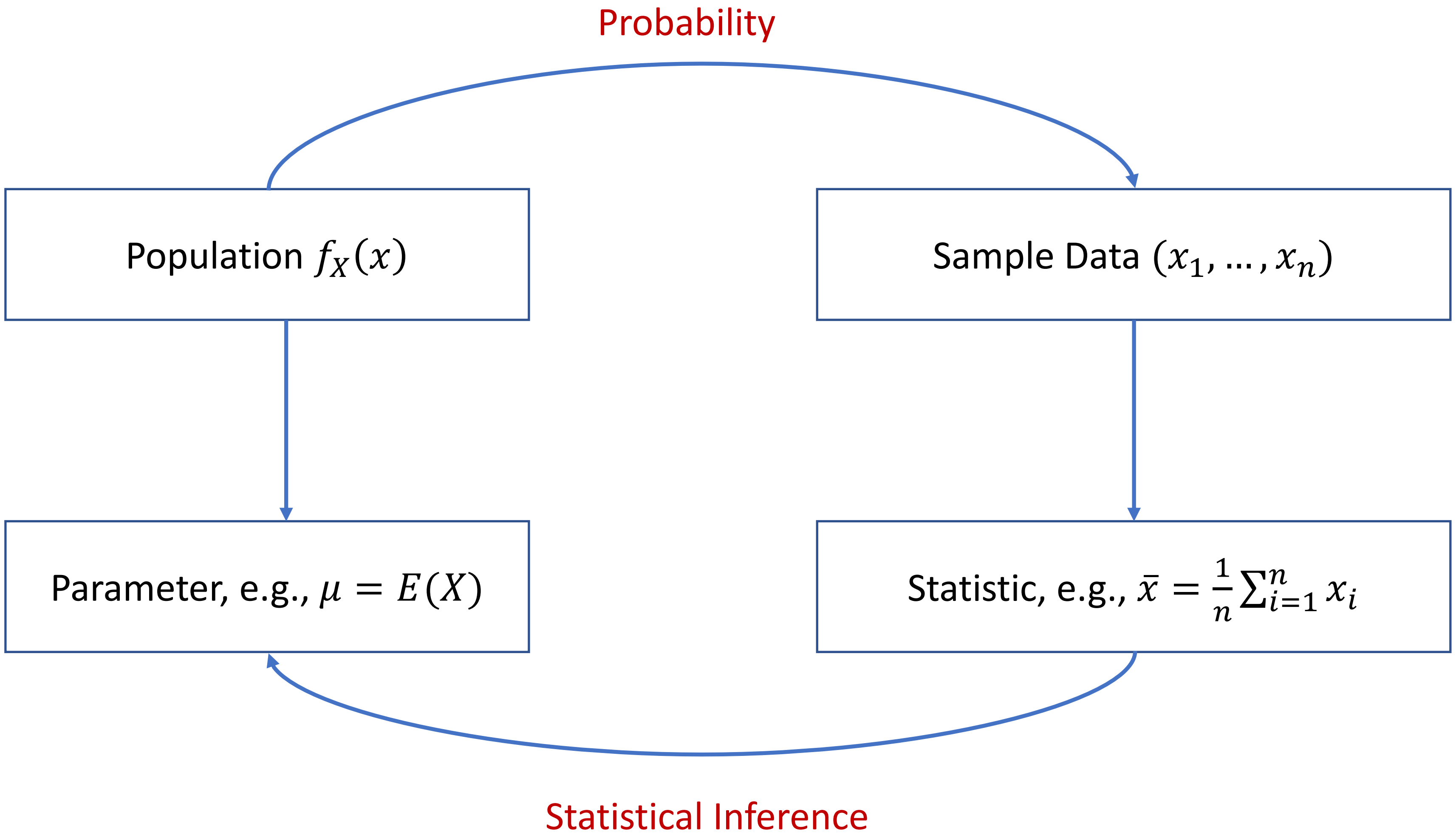
3. $\mathbf{b}$ has a variance-covariance matrix: $V(\mathbf{b}) = \sigma_\epsilon^2(\mathbf{X}'\mathbf{X})^{-1}$.

4. $\mathbf{b}$ has a normal distribution, *if* $\mathbf{y}$ is normally distributed.

Statistical inference

- inference for individual coefficients: t-tests and confidence intervals
- inference for several coefficients: F-tests
- general linear hypotheses

What is the Statistical Inference?



Statistical inference for individual coefficients

- Vector of coefficients $\mathbf{b} = [B_0, B_1, \ldots, B_k]'$

$$\mathbf{b} \sim N_{k+1}(\boldsymbol{\beta}, \sigma_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1}).$$

- Individual coefficient:

$$\mathrm{B}_j \sim N\left(\beta_j, \sigma_\epsilon^2 \mathrm{v}_{jj}\right) \qquad \text{or} \qquad \frac{B_j - \beta_j}{\sigma_\epsilon \sqrt{\mathrm{v}_{jj}}} \sim N(0,1)$$

where v_{jj} is the j-th diagonal entry of $(\mathbf{X}'\mathbf{X})^{-1}$.

Statistical inference for individual coefficients

- For testing $H_0\colon \beta_j = \beta_j^{(0)}$ (e.g., $H_0\colon \beta_j = 1$ or any other value), we could use the test statistic:

$$\mathrm{Z} = \frac{B_j - \beta_j^{(0)}}{\sigma_\epsilon \sqrt{\mathrm{v}_{jj}}}.$$

We assume σ_ϵ^2 would be known here.

- If H_0 is true (i.e., under H_0), $\mathrm{Z} \sim \mathrm{N}(0,1)$.

What is the problem here?

Statistical inference for individual coefficients

1×n n×1   n×1 n×k k×1

- σ_ϵ^2 is estimated by $S_E^2 = \frac{\mathbf{e}'\mathbf{e}}{n-(k+1)}$, where $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{b}$ is the vector of residuals.
- In the variance $V(\mathbf{b}) = \sigma_\epsilon^2(\mathbf{X}'\mathbf{X})^{-1}$, we simply replace σ_ϵ^2 with S_E^2.

k×n n×k   k×k

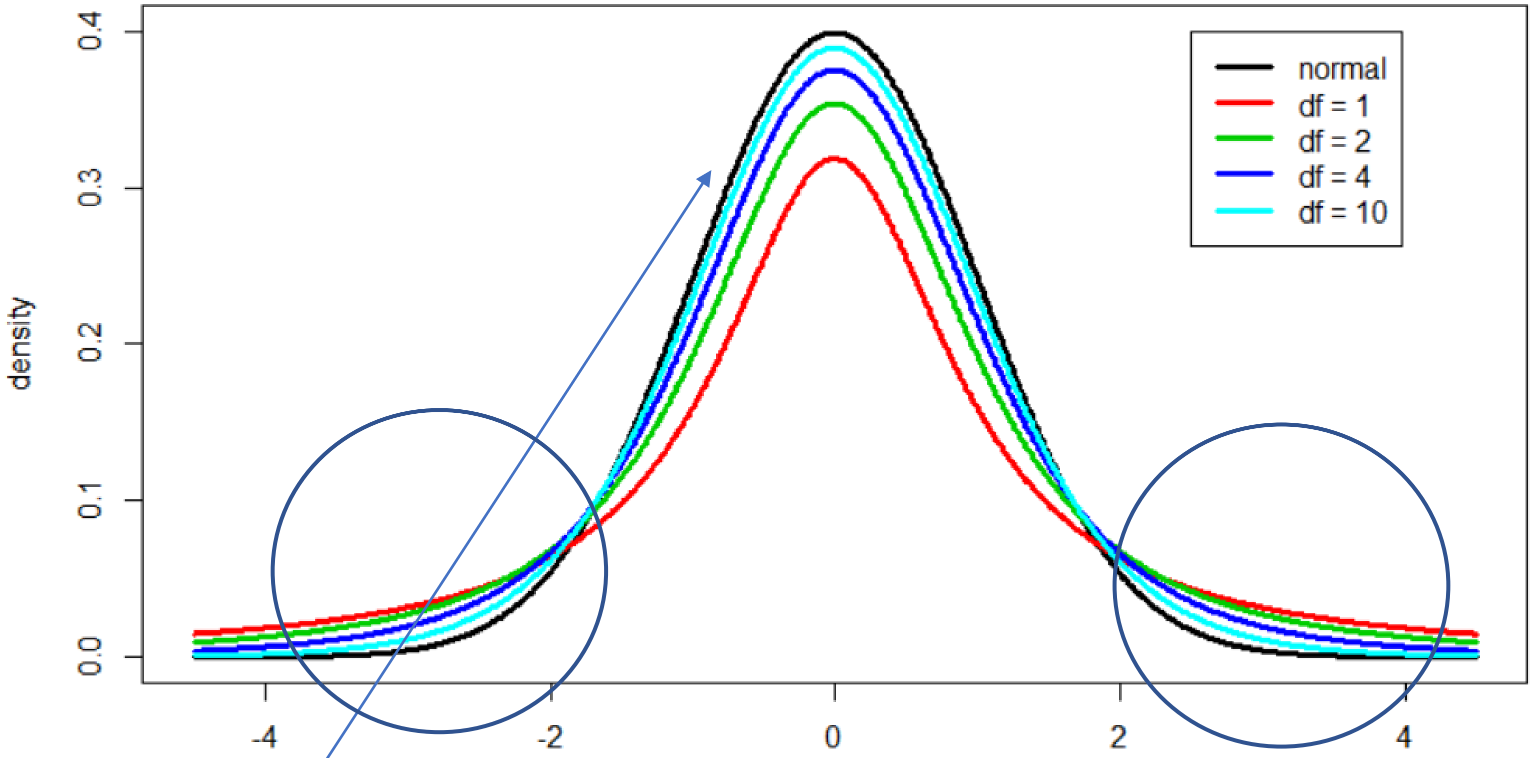
- The estimator of variance-covariance matrix is $\hat{V}(\mathbf{b}) = S_E^2(\mathbf{X}'\mathbf{X})^{-1}$.
- The estimator of standard error is $SE(B_j) = S_E\sqrt{\mathbf{v}_{jj}}$, where $\mathbf{v}_{jj}$ is the j-th diagonal entry of $(\mathbf{X}'\mathbf{X})^{-1}$.
- To test $H_0: \beta_j = \beta_j^{(0)}$, we can use the test statistic

$$t = \frac{B_j - \beta_j^{(0)}}{SE(B_j)} = \frac{B_j - \beta_j^{(0)}}{S_E\sqrt{\mathrm{v}_{jj}}}.$$

- If H_0 is true, then $t \sim t_{n-(k+1)}$.

Student's t-distribution t_n (Generalization of the Standard Normal distribution)

William Gosset (“Student”)



For large df, t-distribution becomes the standard normal $N(0,1)$.

Example: Duncan data

- Dataset on prestige of 45 occupations, to be explained by education and income.

```
> head(duncan,5)
           type income education prestige
accountant prof     62        86       82
pilot      prof     72        76       83
architect  prof     75        92       90
author     prof     55        90       76
chemist    prof     64        86       90
```

- Linear model:

$$\text{prestige}_i = \beta_0 + \beta_1 \text{education}_i + \beta_2 \text{income}_i + \epsilon_i, \text{ for } i = 1, \dots, 45 .$$

Example: Duncan data

```
> Duncanreg <- lm(prestige ~ education + income, data = duncan)
> summary(Duncanreg)

Call:
lm(formula = prestige ~ education + income, data = duncan)

Residuals:
    Min      1Q  Median      3Q     Max 
-29.538  -6.417   0.655   6.605  34.641 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -6.06466    4.27194  -1.420    0.163    
education    0.54583    0.09825   5.555 1.73e-06 ***
income       0.59873    0.11967   5.003 1.05e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.37 on 42 degrees of freedom
Multiple R-squared:  0.8282,	Adjusted R-squared:   0.82 
F-statistic: 101.2 on 2 and 42 DF,  p-value: < 2.2e-16
```

Remember:

R always reports two-tailed P-value for the t-test, with $\beta_j^{(0)} = 0$.

t-test for individual slope, two-sided H_a

- Recall the following steps in hypothesis testing for the slope of education (i.e., β_1):

1. Define the hypothesis test: education is not related to prestige (keeping income constant) vs education is related to prestige (keeping income constant)

$$H_0 : \beta_1 = 0 \text{ vs } H_1 : \beta_1 \neq 0$$

2. Test statistic:

The testing value comes here.

$$t = \frac{B_1 - 0}{SE(B_1)} \qquad \frac{|\hat{\beta}_1 - \beta_1^{H_0}|}{se(\hat{\beta}_1)}$$

3. If H_0 is true, then $t \sim t_{42}$ $\quad$ ($n = 45, k = 2$, therefore df $= 45 - (2+1) = 42$).

4. If H_a is true, then t tends to smaller (if $\beta_1 < 0$) or larger (if $\beta_1 > 0$) values than prescribed by t_{42} distribution.

Example: t-test for individual slope, two-sided H_a

5. Two-tailed p-value is needed :

$$P \ = \ 2 \times \ P(t_{42} \geq |t_{out}|)$$

6. The outcome of test statistic (read from **R** output):

$$t \ = \ \frac{0.546-0}{0.0983} = \ 5.555$$

7. p-value:

$$P \ = \ 2 \times \ P(t_{42} \geq |5.555|) = 2 \cdot 8.65 \cdot 10^{-7} = 1.73 \cdot 10^{-6}.$$

```
pt(5.555, 42, lower.tail = FALSE)
```

Conclusion: $P \ < \alpha = 0.05$, therefore, reject H_0. It is shown that education level required for jobs is related to prestige (keeping income constant).

Example: t-test for individual slope, $H_0: \beta = \beta^{(0)}$

- Suppose a test for $\beta_j^{(0)} \neq 0$.
- R cannot be directly used, unless we use some trick.
- Imagine that the value 0.5 has some special meaning in the education example, and we ask if β_1 might be equal to 0.5 (given the data).

1. Define the hypothesis test:

$$H_0: \beta_1 = 0.5 \quad \text{vs} \quad H_a: \beta_1 \neq 0.5.$$

2. Test statistic (the same as for two-sided test):

$$t = \frac{B_1 - 0.5}{SE(B_1)}$$

3. If H_0 is true, then $t \sim t_{42}$.

Example: t-test for individual slope, $H_0: \beta = \beta^{(0)}$

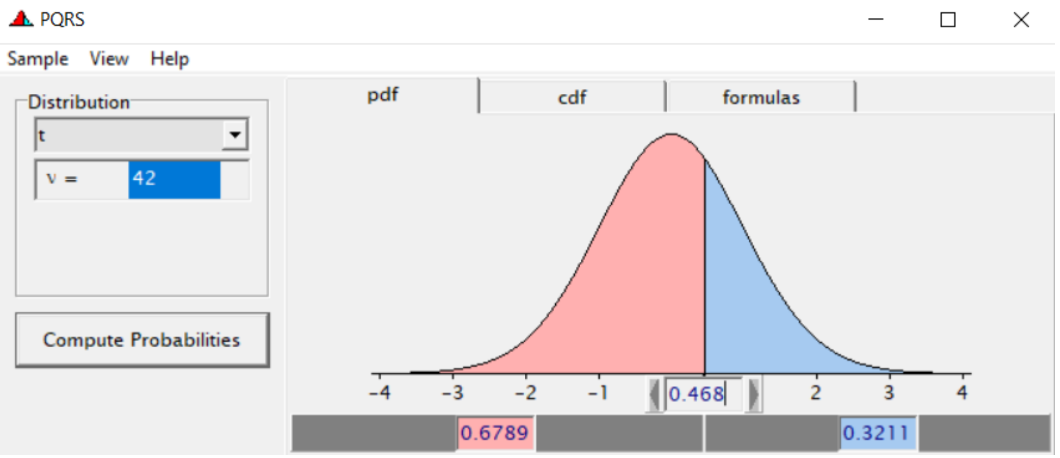
4. If H_a is true, t tends to larger values than prescribed by t_{42} distribution

5. Right-tailed p-value is needed: $P = P(t_{42} > |t|)$

6. The outcome of test statistic: $t = \frac{0.54 - 0.5}{0.0983} = 0.468.$

7. $\mathrm{P} = 2 \times \mathrm{P}(\mathrm{t}_{42} \geq |0.468|) = 2 \times 0.321 = 0.64$

```
pt(0.468, 42, lower.tail = FALSE)
```



Conclusion: $P > 0.05$, do not reject H_0. No evidence is found that the slope deviates from 0.5 (keeping income constant)

Example: t-test for individual slope, one-sided H_a

- Suppose we would like to test if the relationship is positive. In this case, it makes sense to test with a right-sided H_a.The steps are almost the same, with a small difference.

1. Define the hypothesis test:

Education is not related to prestige (keeping income constant) vs education is positively related to prestige (keeping income constant).

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 > 0$$

2. Test statistic: $t = \frac{B_1 - 0}{SE(B_1)}$ (the same as for two-sided test).

3. If H_0 is true, then $t \sim t_{42}$.

4. If H_a is true, t tends to larger values than prescribed by t_{42} distribution

Example: t-test for individual slope, one-sided H_a

5. Right-tailed p-value is needed: $P = P(t_{42} > t)$

6. The outcome of test statistic: $t = \frac{0.54 - 0}{0.0983}$ = 5.555.

7. p-value: $P = P(t_{42} \geq 5.555) = 8.65 \cdot 10^{-7} \cdot$

Conclusion: $P < \alpha = 0.05$, therefore, reject H_0. Thus, education is positively related to prestige (keeping income constant).

Note:

- here we could take the half of the P-value as reported by **R** (i.e., the two-sided p-value).

- Can two-tailed P-value, as reported by **R**, always be halved for one-sided H_a? (No, why?)

Example: t-test for individual slope, one-sided H_a with $\beta^0 \neq 0$

- Suppose we would like to test if Define the hypothesis test:

$$H_0: \beta_1 = 1$$

$$H_1: \beta_1 > 1$$

2. Test statistic: $t = \frac{B_1 - 1}{SE(B_1)}$. $> t_\alpha$ reject

3. If H_0 is true, then $t \sim t_{42}$. $P(t_{42} > \frac{B_1 - 1}{se(B_1)}) < \alpha$

4. If H_a is true, t tends to larger values than prescribed by t_{42} distribution

5. Right-tailed p-value is needed: $P = P(t_{42} > t)$

6. The outcome of test statistic: $t = \frac{0.54 - 1}{0.0983} = -4.679.$

7. p-value: $P = P(t_{42} \geq -4.679) = 0.9999$ ·

Conclusion: $P > \alpha = 0.05$, therefore, failed to reject H_0.

Confidence interval for slope

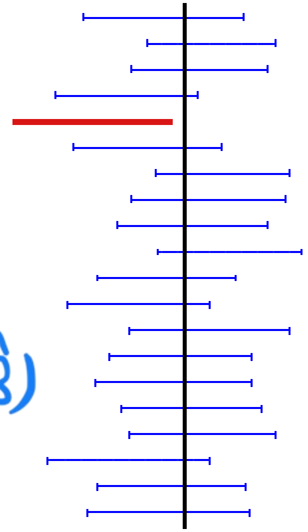
$$\frac{|\hat{\beta}-\beta_0|}{se(\hat{\beta})} \leq t_\alpha$$

- We can also use $100(1-\alpha)\%$ confidence interval

$$-t_\alpha se(\hat{\beta}) \leq \hat{\beta}-\beta_0 \leq t_\alpha se(\hat{\beta})$$

$$CI(\beta_j) = \mathrm{B}_j \pm t_{\alpha/2\,;\,n-(k+1)} SE(\mathrm{B}_j)$$

$$\hat{\beta}-t_\alpha se(\hat{\beta}) \leq \beta_0 \leq \hat{\beta}+t_\alpha se(\hat{\beta})$$



```
> confint(Duncanreg)
                 2.5 %    97.5 %
(Intercept) -14.6857892 2.5564634
education     0.3475521 0.7441158
income        0.3572343 0.8402313
```

- The CI's do not contain the value 0, which means we can reject the two-sided test against 0.

Example: confidence interval for slope

```
> coef(summary(Duncanreg))
              Estimate Std. Error   t value     Pr(>|t|)
(Intercept) -6.0646629 4.27194117 -1.419650 1.630896e-01
education    0.5458339 0.09825264  5.555412 1.727192e-06
income       0.5987328 0.11966735  5.003310 1.053184e-05
```

- $100(1-\alpha)\%$ level confidence interval for slope is defined as:

$$CI(\beta_1) = B_1 \pm t_{\alpha/2;\,45-3}\,SE(B_1) =$$

$$0.546 \pm t_{42;\,0.025} \times 0.0983 \ =$$

$$= 0.546 \pm 2.018 \times 0.0983 =$$

$$= (0.348; 0.744)$$

Statistical inference for several coefficients: All-slopes

- Multiple regression model for response Y_i and k regressors $x_1, \ldots, x_k$:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i, \text{ for } i = 1, \ldots, n$$

- Global or "omnibus" test that all regressors are unimportant.

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_1: \text{at least one slope is not zero / at least one x has predictive value}$$

- In this case, the F-test statistic is used:

$$F = \frac{RegMS}{RMS} = \frac{RegSS/k}{RSS/(n-(k+1))}$$

- Recall, $RegSS = TSS - RSS$, i.e., difference between the residual sum of squares of the null model (i.e., intercept-only model) and current model.

Statistical inference for several coefficients: All-slopes

- $F = \frac{RegMS}{RMS}$ is a ratio of two Mean Squares:
  - Denominator: Residual Mean Square RMS is an estimator of the error variance σ_ϵ^2.
  - Numerator: Regression Mean Square $RegMS$ is also an estimator of σ_ϵ^2, **but only if $\boldsymbol{H_0}$ is true!**
- Hence, under H_0, the ratio $RegMS$ / RMS is close to 1.
- Under H_a, $RegMS$ tends to be larger than σ_ϵ^2, so the ratio tends to be larger than 1.

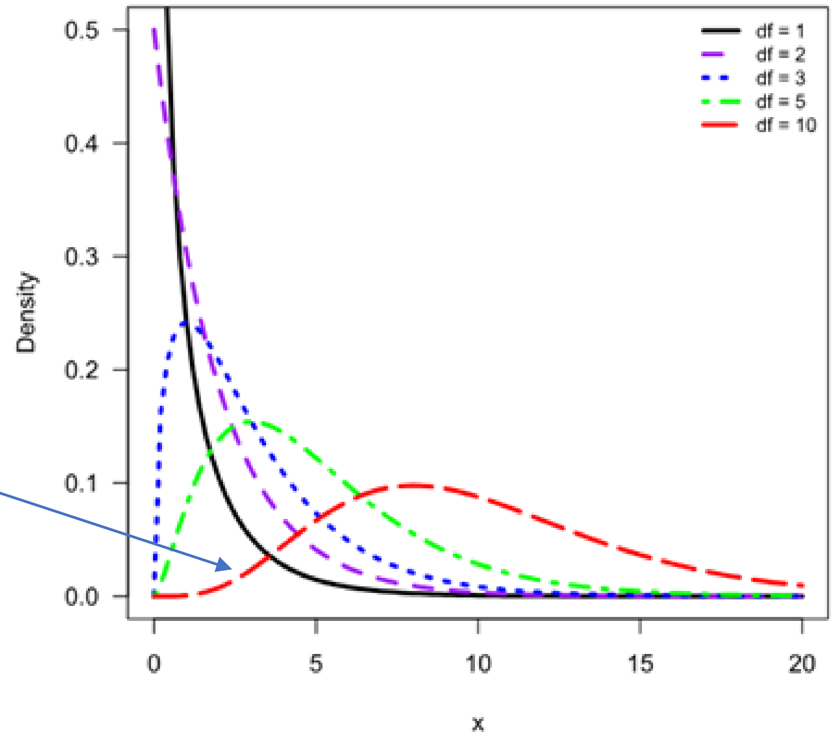
- If H_0 is true (i.e., under H_0), $F \sim F_{k;\, n-(k+1)}$
- Reject H_0 for large values of F, right-sided P-value and rejection region.

Chi-squared distribution χ_k^2

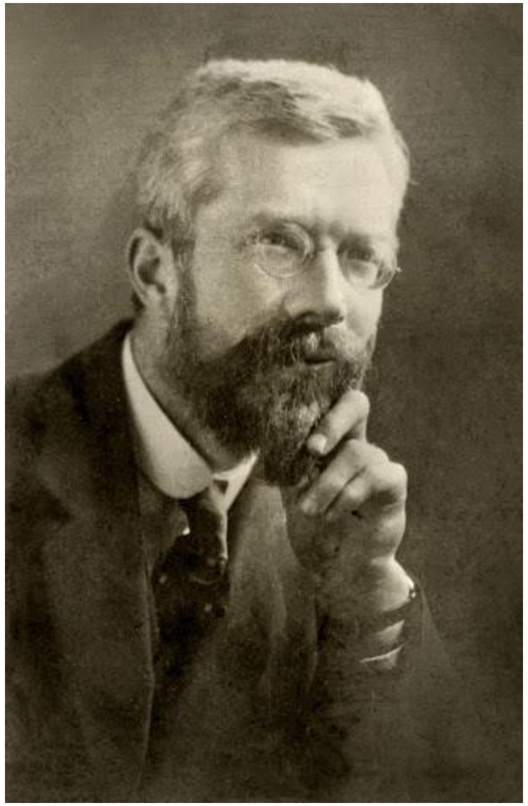
- Suppose $Z_1, \dots, Z_m$ are independent, standard normal random variables, i.e., $Z_i \sim N(0,1)$
- Sum of their squares follows a χ_m^2 distribution, with $\cancel{n}$ m degrees of freedom.

$$X^2 = \sum_{i=1}^{m} Z_i^2 \sim \chi_m^2$$

- The mean $E(X^2) = m$ (i.e., df)



F-distribution



Ronald Fisher

- Suppose $X_1^2 \sim \chi^2_{\mathrm{df}_1}$ and $X_2^2 \sim \chi^2_{\mathrm{df}_2}$ are two independent chi-square distributed variables, with degrees of freedom df_1 and df_2, respectively.

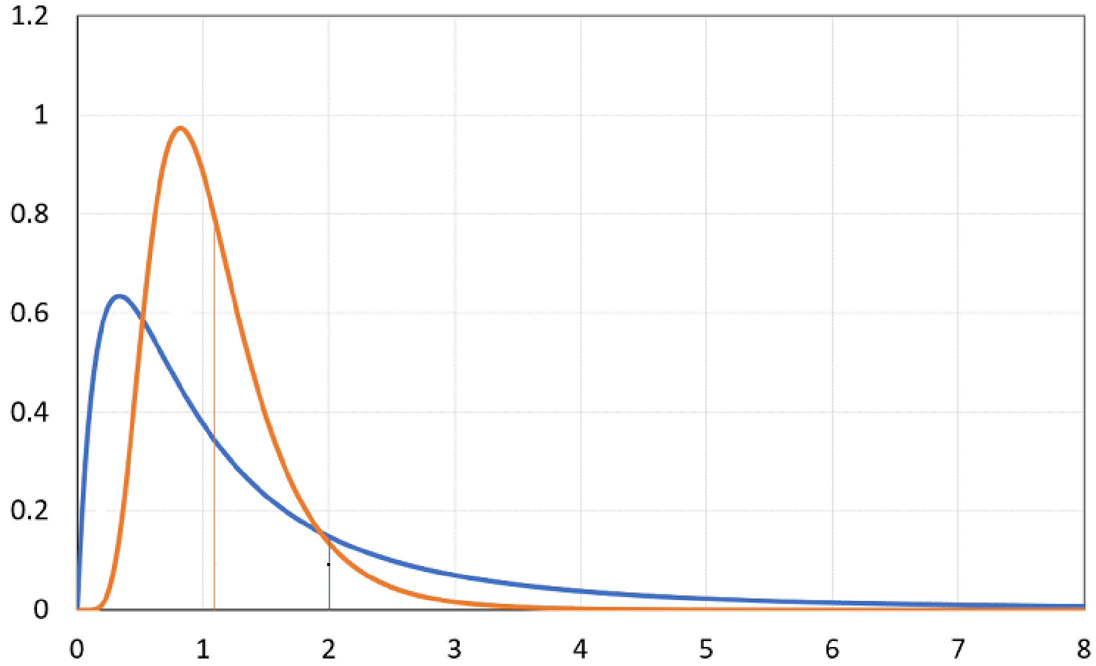
- F-distribution is obtained by taking the ratio

$$F \equiv \frac{X_1^2/\mathrm{df}_1}{X_2^2/\mathrm{df}_2} \sim F_{\mathrm{df}_1;\,\mathrm{df}_2}.$$

- F-distribution has two degrees of freedom: numerator df df_1 and denominator df_2.

F-distribution: Examples

- Blue line: $\mathrm{df}_1 = 4, \mathrm{df}_2 = 4.$
- Orange line: $\mathrm{df}_1 = 20, \mathrm{df}_2 = 20.$



- The mean $E(F) = \frac{df_2}{df_2 - 2}$, for $df_2 > 2$.

- If $t \sim \mathrm{t_{df}}$ then $t^2 \sim F_{1;\mathrm{df}}$.
- For $q = 1$ F-test is equivalent to t-test.

Statistical inference for several coefficients: All-slopes

- Analysis of variance table or ANOVA table shows construction of F (a reminder)

Source	Sum of Squares	df	Mean Square	F
Regression	$RegSS$	k	$\frac{RegSS}{k}$	$\frac{RegMS}{RMS}$
Residual	RSS	$n-(k+1)$	$\frac{RSS}{n-(k+1)}$	
Total	TSS	$n-1$		

- k is the number of regressors in the model.

Example: Duncan data

```
> Duncanreg <- lm(prestige ~ education + income, data = duncan)
> summary(Duncanreg)

Call:
lm(formula = prestige ~ education + income, data = duncan)

Residuals:
    Min      1Q  Median      3Q     Max 
-29.538  -6.417   0.655   6.605  34.641 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -6.06466    4.27194  -1.420    0.163    
education    0.54583    0.09825   5.555 1.73e-06 ***
income       0.59873    0.11967   5.003 1.05e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.37 on 42 degrees of freedom
Multiple R-squared:  0.8282,    Adjusted R-squared:   0.82 
F-statistic: 101.2 on 2 and 42 DF,  p-value: < 2.2e-16
```

What is your conclusion?

Example: Duncan data

```
> anova(Duncanreg)
Analysis of Variance Table

Response: prestige
          Df Sum Sq Mean Sq F value    Pr(>F)
education  1  31707   31707 177.399 < 2.2e-16 ***
income     1   4474    4474  25.033 1.053e-05 ***
Residuals 42   7507     179
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> (RegSS <- sum((fitted.values(Duncanreg) - mean(fitted.values(Duncanreg)))^2))
[1] 36180.95
> (RSS <- deviance(Duncanreg))  #Easy way to obtain the RSS
[1] 7506.699
> (Fstat <- (RegSS/(Duncanreg$rank - 1)) / (RSS/Duncanreg$df.residual))
[1] 101.2162
```

R reports the sums of squares of education and income.

$$\underbrace{(y_i-\bar{y})^2}_{TSS} = \underbrace{(\hat{y}_i-\bar{y})^2}_{RSS} + \underbrace{(y_i-\hat{y}_i)^2}_{ESS} \qquad 2(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$$

Example: Duncan data

- As for the t-test, we have the following steps for the F-test.

1. Hypothesis test:

$$H_0: \beta_1 = \beta_2 = 0$$

$$H_1: \text{at least one is not zero.}$$

2. Test statistic:

$$F = \frac{RegSS/k}{RSS/(n-(k+1))}.$$

3. If H_0 is true $F \sim F_{2;42}$.

4. If H_a is true, F tends to larger values than prescribed by $F_{2;42}$ distribution.

Example: Duncan data

5. Right-tailed p-value: $P = P(F_{2;42} \geq F)$

6. The outcome of test statistic

$$F = \frac{RegSS/2}{RSS/42} = \frac{36181/2}{7507/42} = 101.2.$$

7. P-value: $P = P\left(F_{2;42} \geq 101.2\right) = 8.76 \times 10^{-16}$.

Conclusion: $P < 0.05$, so reject H_0. Therefore, education and/or income are related to prestige.

- To calculate the p-value use: `pf(101.2, 2, 42, lower.tail =FALSE)`

Statistical inference for several coefficients: Subset of Slopes

- Inference on groups of coefficients may be needed because
  - least-squares estimators are often correlated (off-diagonal elements of $V(\mathbf{b})$ are non-zero).
  - interest in related set of coefficients, like in ANOVA.

- Suppose we would like to test if a subset of slopes are 0, instead of all slopes

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_q = 0$$

$$H_1: at\ least\ one\ is\ not\ zero$$

- For notational convenience, let's focus on the first q regressors, but any subset of β_i's may be tested.

Hypothesis Test: Subset of Slopes

- F-test is constructed by fitting two nested models:

Full (or initial) model FM:

$$Y = \beta_0 + \beta_1 x_1 + \cdots \beta_q x_q + \beta_{q+1} x_{q+1} + \cdots + \beta_k x_k + \epsilon.$$

Reduced model RM:

$$Y = \beta_0 + 0x_1 + \cdots + 0x_q + \beta_{q+1} x_1 + \cdots + \beta_k x_k + \epsilon = \beta_0 + \beta_{q+1} x_1 + \cdots + \beta_k x_k + \epsilon$$

FM and RM give residual Sum of Squares RSS_1 and RSS_0, respectively.

Statistical inference for several coefficients: Subset of Slopes

- We have $\overset{\text{FM}}{\underline{RSS = \mathbf{e}'\mathbf{e}}}$ residual sum of squares of FM and $\overset{\text{RM}}{\underline{RSS_0 = \mathbf{e}_0'\mathbf{e}_0}}$ residuals sum of squares of RM.
- The F-ratio is defined as

$$F_0 = \frac{(RSS_0 - RSS)/q}{RSS/\big(n - (k+1)\big)}$$

- Under H_0 , $F_0 \sim F_{q;n-(k+1)}$.

- Is F-ratio always positive? Why?

- The following holds: $RSS_0 - RSS = RegSS - RegSS_0$, i.e., "Any increase in residual sum of squares, is decrease in regression sum of squares“.

- Therefore, we can write $F = \frac{(RegSS - RegSS_0)/q}{RSS_1/(n-(k+1))}$. $\frac{(RegSS - RegSS_0)/q}{RSS/[n-(k+1)]}$

Example: Duncan data

- Suppose we would like to test $H_0: \beta_1 = 0$ (i.e., education has no association with prestige).

```
> DuncanregFM <- lm(prestige ~ education + income, data = duncan)
> DuncanregRM <- lm(prestige ~  income, data = duncan)
> (RSS <- deviance(DuncanregFM))
[1] 7506.699
> (RSS0 <- deviance(DuncanregRM))
[1] 13022.8
> q <- 1
> (Fstat <- ((RSS0 - RSS)/q) / (RSS/DuncanregFM$df.residual))
[1] 30.8626
> (pval <- pf(Fstat, q, DuncanregFM$df.residual, lower.tail = FALSE))
[1] 1.727192e-06
```

- Remember that, if $t \sim \mathrm{t}_{\mathrm{df}}$ then $t^2 \sim F_{1;\mathrm{df}}$.
- For $q = 1$ F-test is equivalent to t-test.

Example: Duncan data

- Another approach in **R**

```
> anova(DuncanregFM, DuncanregRM)
Analysis of Variance Table

Model 1: prestige ~ education + income
Model 2: prestige ~ income
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     42  7506.7
2     43 13022.8 -1   -5516.1 30.863 1.727e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Statistical inference for several coefficients: Subset of Slopes

- Let $\mathbf{b}_1 = \left[\mathbf{B}_1, \dots, \mathbf{B}_q\right]'$ be LS coefficients of interest from $\mathbf{b}$ and $\mathbf{V}_{11}$ be the corresponding submatrix of $(\mathbf{X}'\mathbf{X})^{-1}$.

- We can show that $RSS_0 - RSS = \mathbf{b}_1'\mathbf{V}_{11}^{-1}\mathbf{b}_1$

$$F_0 = \frac{(RSS_0 - RSS)/q}{RSS/\left(n-(k+1)\right)} = \frac{\mathbf{b}_1'\mathbf{V}_{11}^{-1}\mathbf{b}_1}{qS_E^2}$$

- Test a general hypothesis $H_0\colon \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{(0)}$, where $\boldsymbol{\beta}_1 = \left[\beta_1, \beta_2, \dots, \beta_q\right]'$ and $\boldsymbol{\beta}_1^{(0)}$ not necessarily $\mathbf{0}$.

$$F_0 = \frac{\left(\mathbf{b}_1 - \boldsymbol{\beta}_1^{(0)}\right)' \mathbf{V}_{11}^{-1} \left(\mathbf{b}_1 - \boldsymbol{\beta}_1^{(0)}\right)}{qS_E^2} \sim F_{q;n-(k+1)}$$

Statistical inference

- inference for individual coefficients: t-tests and confidence intervals
- inference for several coefficients: F-tests
- general linear hypotheses

General linear hypotheses

- Consider the following linear hypothesis: $H_0 \colon \mathbf{L}_{q\times(k+1)}\boldsymbol{\beta}_{(k+1)\times 1} = \mathbf{c}_{q\times 1}$
- The hypothesis matrix $\mathbf{L}$ is full row rank $q \le k+1$.
- The F-statistic is defined as: $F_0 = \dfrac{(\mathbf{L}\boldsymbol{\beta}-\mathbf{c})'\left[\mathbf{L}\left(\mathbf{X}'\mathbf{X}\right)^{-\mathbf{1}}\mathbf{L}'\right]^{-1}(\mathbf{L}\boldsymbol{\beta}-\mathbf{c})}{qS_E^2} \sim F_{q,n-(k+1)}$ (under H_0), because
  - $\mathbf{b} \sim N_{k+1}(\boldsymbol{\beta}, \sigma_\epsilon^2(\mathbf{X}'\mathbf{X})^{-1})$
  - $\mathbf{Lb} \sim N_q(\mathbf{L}\boldsymbol{\beta}, \sigma_\epsilon^2\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')$
  - $(\mathbf{L}\boldsymbol{\beta} - \mathbf{c})'\left[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{L}'\right]^{-1}(\mathbf{L}\boldsymbol{\beta} - \mathbf{c})/\sigma_\epsilon^2 \sim \chi_q^2$, under H_0

General linear hypotheses

- Example (Practical exercise)
- Consider the hypothesis:

$$H_0: \beta_1 = \beta_2 = 0$$

- We take $\mathbf{L} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $\mathbf{c} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

$$\mathbf{L}_{q\times(k+1)}\boldsymbol{\beta}_{(k+1)\times 1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

$$= \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- Consider

$$H_0: \beta_1 - \beta_2 = 0$$

- Define $\mathbf{L} =?$ and $\mathbf{c} =?$

$$\begin{pmatrix} 1 & -1 \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = 0$$

Predicting new y-values

- Forecasting the future response values:
  - E.g., predicting the prestige values based on education and income.

- Two possible interpretation of prediction based on a given x.
  - The estimate of the mean (average) prestige $\mu_y = E(y)$ at specific values of education and income:

$$\hat{\mu}_y = B_0 + B_1 x^*_{n+1\ 1} + \cdots + B_k\, x^*_{n+1\ k}$$

  - Estimated prestige at the specific value of education and income:

$$\hat{Y}_{n+1} = B_0 + B_1 x^*_{n+1\ 1} + \cdots + B_k\, x^*_{n+1\ k}$$

Example: Duncan data

- Suppose, we want to predict the prestige value for a new profession with
  - education $= 92$
  - income $= 68$

```
> xnew <- data.frame(education = 92, income = 68)
> predict(Duncanreg, newdata = xnew)
       1
84.86589
>
> coef(Duncanreg)[1] + coef(Duncanreg)[2]*92 + coef(Duncanreg)[3]*68
(Intercept)
   84.86589
```

- Extrapolation in regression:
  - Be concerned not only about individual predictor but also about the set of values of several predictors together.

Inference for predictions

- Confidence interval for μ_y:

$$\mathrm{CI}(\mu_y) = \hat{\mu}_y \pm t_{\mathrm{df}E;\alpha/2}\,\mathrm{se}(\hat{\mu}_y)$$

$$x^* = [1, x_1^*, \dots, x_k^*]$$

where $\mathrm{df}E$ is the df of the error term and

$$\mathrm{se}(\hat{\mu}_y) = S_E\sqrt{x^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}x^*} = \sqrt{S_E^2(x^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}x^*)}$$

- Prediction interval for individual Y:

$$\mathrm{CI}(\hat{Y}) = \hat{Y} \pm t_{\mathrm{df}E;\alpha/2}\,\mathrm{se}(\hat{Y})$$

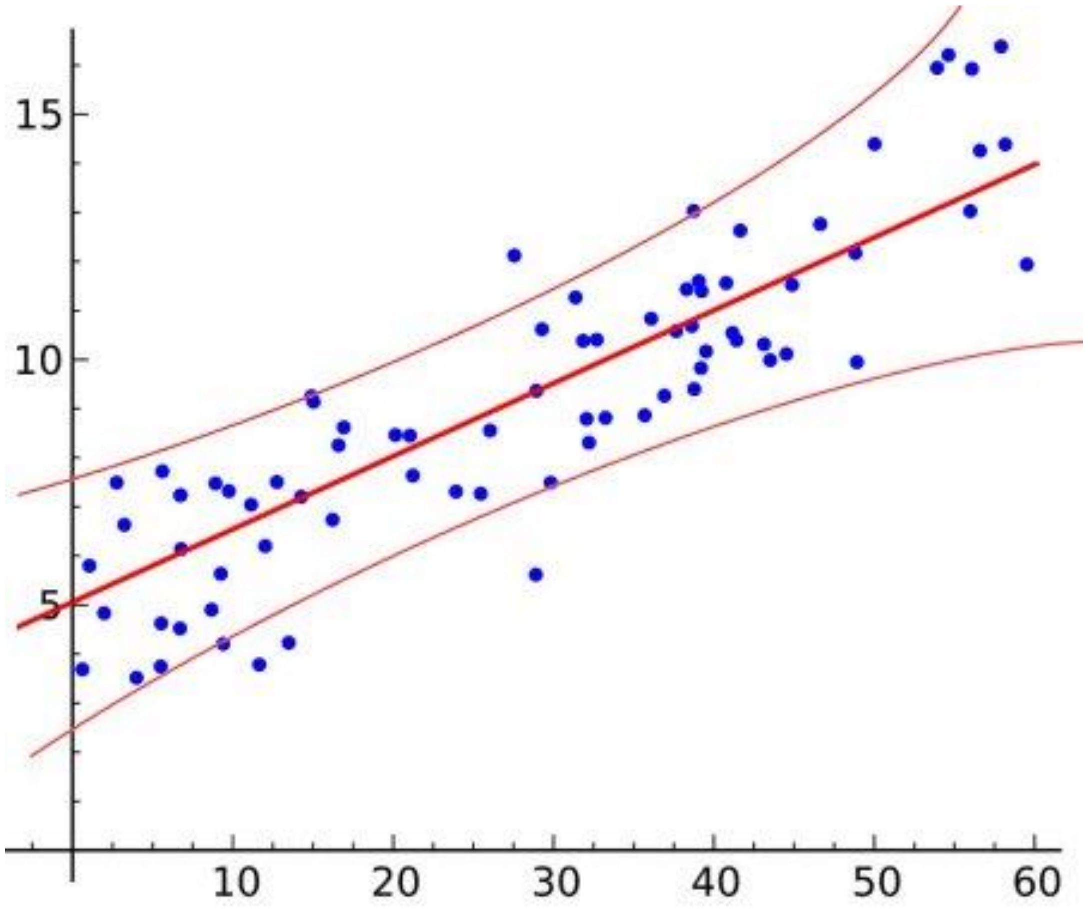
Which one is larger and why?

where $\mathrm{df}E$ is the df of the error term and

$$\mathrm{se}(\hat{Y}) = S_E\sqrt{x^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}x^* + 1} = \sqrt{S_E^2(x^{*\prime}(\mathbf{X}'\mathbf{X})^{-1}x^*) + S_E^2}$$

$$= \sqrt{\mathrm{se}(\hat{\mu}_y)^2 + S_E^2}$$

Confidence vs prediction interval



- A CI gives a range for $E(y)$ and a PI gives a range for y.
- A PI is wider than a CI because it includes a wider range of values.
- A PI predicts an individual value, whereas a CI predicts the mean value.
- A PI focuses on the future values, whereas a CI focuses on past values.

Example: Duncan data

```
> #Confidence interval
> pr1 <- predict(Duncanreg, newdata = xnew, interval = "confidence",se.fit=TRUE)
> pr1$fit
       fit      lwr      upr
1 84.86589 78.10926 91.62252
>
> #Prediction interval
> pr2 <- predict(Duncanreg, newdata = xnew, interval = "prediction",se.fit=TRUE)
> pr2$fit
       fit      lwr      upr
1 84.86589 57.05292 112.6789
>
> #Manual calculations
> X <- model.matrix(Duncanreg)
> #New x_star
> newX <- as.vector(c(1, 92, 68))
> (se_mu <- summary(Duncanreg)$sigma * sqrt(t(newX)%*%solve(t(X)%*%X)%*%(newX)))
         [,1]
[1,] 3.348046
> pr1$se.fit
[1] 3.348046
> c(pr1$fit[1]  - qt(0.025, pr1$df, lower.tail = FALSE)*se_mu,
+   pr1$fit[1] + qt(0.025, pr1$df, lower.tail = FALSE)*se_mu)
[1] 78.10926 91.62252
>
> se_ind <- sqrt(pr1$se.fit^2 + pr1$residual.scale^2)
> c(pr1$fit[1]  - qt(0.025, pr1$df, lower.tail = FALSE)*se_ind,
+   pr1$fit[1] + qt(0.025, pr1$df, lower.tail = FALSE)*se_ind)
[1]  57.05292 112.67886
```

- Specify the ‘`interval`’ argument for CI or PI.

- `predict()` function provides the standard error of the predicted means.