To Do

Read Sections 6.1 – 6.2.

Assignment 4 is due Friday November 25.

Last Class

(1) Distribution of $\tilde{\beta}$ the Maximum Likelihood Estimator of the Slope (with Proof)

(2) Distribution of S_e^2 the Estimator of σ^2 (no Proof)

Today's Class

(1) Confidence Interval for Slope β and Testing H_0: $\beta = \beta_0$

(2) Test of No Relationship between Response and Explanatory Variates

(3) Confidence interval for the mean response $\mu(x) = \alpha + \beta x$

(4) Prediction Interval for an Individual Response Y

Theorem from Last Day

Since

$$\frac{\tilde{\beta}-\beta}{\sigma / \sqrt{S_{XX}}} \sim G(0,1) \quad \text{and} \quad \frac{(n-2)S_e^2}{\sigma^2} \sim \chi^2(n-2)$$

independently then

$$\frac{\tilde{\beta}-\beta}{S_e / \sqrt{S_{XX}}} \sim t(n-2)$$

This pivotal quantity can be used to construct confidence intervals and test hypotheses for *β*.

Inferences for the Slope *β* - Summary

A 100p% confidence interval for *β* is given by

$$\hat{\beta} \pm a s_e / \sqrt{S_{XX}}$$

where P(*T* ≤ *a*) = (1 + *p*)/2 and *T* ∽ *t*(*n*-2).

For testing H_0: $\beta = \beta_0$ the *p*-value is

$$p - value = 2\left[1 - P\left(T \le \frac{\left|\hat{\beta} - \beta_0\right|}{s_e \big/ \sqrt{S_{XX}}}\right)\right]$$

where *T* ∽ *t*(*n*-2).

Hypothesis of No Relationship

Since $\mu(x) = \alpha + \beta x$, a test of

$$H_0\text{: } \beta = 0$$

is a test of the hypothesis that the mean $\mu(x)$ does not depend on x.

This hypothesis is usually referred to as "the hypothesis of no relationship" between the variates Y and x.

STAT 230 and 231 Final Grades

$\overline{x} = 76.7333 \qquad \overline{y} = 72.2333$

$S_{XX} = 5135.8667 \qquad S_{XY} = 5106.8667 \qquad S_{XY} = 7585.3667$

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{5106.8667}{5135.8667} = 0.9944$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\,\overline{x} = 72.2333 - \left(\frac{5106.8667}{5135.8667}\right)(76.7333) = 4.0667$$

$$s_e = \sqrt{\frac{1}{n-2}\left(S_{YY} - \hat{\beta} S_{XY}\right)}$$

$$= \sqrt{\frac{1}{28}\left[7585.3667 - (0.9944)(5106.8667)\right]} = 9.4630$$

95% Confidence Interval for *β*

Since P(*T* ≤ 2.0484) = (1+0.95)/2 = 0.975 where *T* ∽ *t*(28) a 95% confidence interval for *β* is

$$\hat{\beta} \pm 2.0484 s_e / \sqrt{S_{XX}}$$

$$= 0.9944 \pm 2.0484(9.4630)/\sqrt{5135.8667}$$

or [0.7239, 1.2648].

95% Confidence Interval for β

Since the 95% confidence interval for β, [0.7239, 1.2648] does not contain the value $\beta = 0$, the *p*-value for testing H_0: $\beta = 0$ is smaller than 0.05.

Therefore there is evidence against the hypothesis of no relationship between STAT 231 final grades and STAT 230 final grades.

p-value for Testing H_0: $\beta = 0$

The actual p-value for testing H_0: $\beta = 0$ is

$$p-value = 2\left[1 - P\left(T \le \frac{\left|\hat{\beta} - 0\right|}{s_e \big/ \sqrt{S_{xx}}}\right)\right] \text{ where } T \sim t(28)$$

$$= 2\left[1 - P\left(T \le \frac{0.9944}{9.4630 \big/ \sqrt{5135.8667}}\right)\right]$$

$$= 2\left[1 - P\left(T \le 7.5304\right)\right] \approx 0$$

There is very strong evidence against the hypothesis of no relationship between STAT 230 final grades and STAT 231 final grades.

H_0: $\beta = 1$

What does the hypothesis H_0: $\beta = 1$ represent?

H_0: $\beta = 1$

The parameter β represents the change in the mean STAT 231 final grade in the study population for a one mark increase in STAT 230 final grade.

The hypothesis H_0: $\beta = 1$ means that we are hypothesizing that in the study population for every one mark increase in STAT 230 final grade there is a one mark increase in the mean STAT 231 final grade.

H_0: $\beta = 1$

Since the 95% confidence interval,

[0.7239,1.2648]

does contain the value $\beta = 1$, the *p*-value for testing H_0: $\beta = 1$ is larger than 0.05 and there is no evidence against the hypothesis H_0: $\beta = 1$.

H_0: $\beta = 1$

The actual p-value for testing H_0: $\beta = 1$ is

$$p-value = 2\left[1-P\left(T \le \frac{\left|\hat{\beta}-1\right|}{s_e \big/ \sqrt{S_{XX}}}\right)\right] \text{ where } T \sim t(28)$$

$$= 2\left[1-P\left(T \le \frac{\left|0.9944-1\right|}{9.4630 \big/ \sqrt{5135.8667}}\right)\right]$$

$$= 2\left[1-P\left(T \le 0.0428\right)\right]$$

Since P(T ≤ 0.2558) = 0.6, *p*-value ≥ 2(1 - 0.6) = 0.8.

Confidence Interval for $\mu(x) = \alpha + \beta x$

Suppose we wanted a confidence interval for the mean STAT 231 final grade for students who obtained a final grade of 75 in STAT 230, that is, we want a confidence interval for $\mu(75) = \alpha + \beta(75)$.

More generally we are often interested in a confidence interval for the mean response $\mu(x) = \alpha + \beta x$ for a specified value of x.

Confidence Interval for $\mu(x) = \alpha + \beta x$

By the Invariance Property of Maximum Likelihood Estimates the maximum likelihood estimator of $\mu(x)$ is the random variable

$$\tilde{\mu}(x) = \tilde{\alpha} + \tilde{\beta}x = \bar{Y} + \tilde{\beta}(x - \bar{x})$$

We need the distribution of $\tilde{\mu}(x)$ **to construct a confidence interval for the mean response $\mu(x) = \alpha + \beta x$.**

Confidence Interval for $\mu(x) = \alpha + \beta x$

It can be shown that (with some effort!)

$$\tilde{\mu}(x) = \tilde{\alpha} + \tilde{\beta}x = \overline{Y} + \tilde{\beta}(x - \overline{x})$$

$$= \sum_{i=1}^{n}\left[\frac{1}{n} + (x - \overline{x})\frac{(x_i - \overline{x})}{S_{XX}}\right]Y_i$$

where $Y_i \backsim G(\alpha + \beta x_i, \sigma)$ for $i = 1,2,\ldots,n$ independently

Distribution of $\tilde{\mu}(x)$

$$\tilde{\mu}(x) \sim G\left(\mu(x), \sigma\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{XX}}} \right)$$

where

$$\tilde{\mu}(x) = \tilde{\alpha} + \tilde{\beta}x$$

and

$$\mu(x) = \alpha + \beta x$$

Distribution of $\tilde{\mu}(x)$

Equivalently

$$\frac{\tilde{\mu}(x) - \mu(x)}{\sigma\sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}} \sim G(0,1)$$

where

$$\tilde{\mu}(x) = \tilde{\alpha} + \tilde{\beta}x$$

and

$$\mu(x) = \alpha + \beta x$$

Distribution of $\tilde{\mu}(x)$

Since we don't know *σ* we use

$$\frac{\tilde{\mu}(x) - \mu(x)}{S_e\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \sim t(n-2)$$

to construct confidence intervals for
***μ*(*x*) = *α* + *βx*.**

Confidence Interval for $\mu(x) = \alpha + \beta x$

A 100*p*% confidence interval for
$\mu(x) = \alpha + \beta x$ = mean response at x is

$$\hat{\mu}(x) \pm a s_e \sqrt{\frac{1}{n} + \frac{\left(x - \bar{x}\right)^2}{S_{xx}}}$$

$$= \hat{\alpha} + \hat{\beta}x \pm a s_e \sqrt{\frac{1}{n} + \frac{\left(x - \bar{x}\right)^2}{S_{xx}}}$$

where $P(T \le a) = (1+p)/2$ and $T \sim t(n - 2)$.

Confidence Interval for the Intercept α

Since $\mu(0) = \alpha + \beta(0) = \alpha$, a 100p% confidence interval for α, is given by

$$\hat{\alpha} \pm a s_e \sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{S_{XX}}}$$

If $\bar{x}$ is large in magnitude (which means the average x_i is large), then the confidence interval for α will be very wide.

This would be disturbing if the value $x = 0$ is a value of interest, but often it is not.

STAT 230/231 Example

Since P($T \le 2.0484$) = 0.975 where $T \sim t(28)$ a 95% confidence interval for the mean STAT 231 final grade for students who obtained a final grade of 75 in STAT 230 is

$$\hat{\alpha} + \hat{\beta}x \pm a s_e \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}$$

$$= -4.0667 + 0.9944(75)$$

$$\pm 2.0484(9.4630)\sqrt{\frac{1}{30} + \frac{(75 - 76.7333)^2}{5135.8667}}$$

$$= 70.51 \pm 3.5699$$

or [66.9,74.1]

Confidence Interval for an Individual Response Y at x

"It's all about me."

Suppose we wanted an interval for Y = the STAT 231 mark for one student who obtained a mark of $x = 75$ in STAT 230.

Confidence Interval for an Individual Response *Y* at *x*

Let *Y* = potential observation for given value of *x*.

Since $$Y \sim G(\alpha + \beta x, \sigma)$$

and $$\tilde{\mu}(x) \sim G\left(\alpha + \beta x, \sigma \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{XX}}} \right)$$

$$Y - \tilde{\mu}(x) \sim G\left(0, \sigma \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{XX}}} \right)$$

Confidence Interval for an Individual Response *Y* at *x*

Equivalently

$$\frac{Y-\tilde{\mu}(x)}{\sigma\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}} \sim G(0,1)$$

Since we don't know *σ* we use

$$\frac{Y-\tilde{\mu}(x)}{S_e\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}} \sim t(n-2)$$

100*p*% Prediction Interval for a Future Response *Y*

The corresponding interval is

$$\hat{\alpha} + \hat{\beta}x \pm s_e \sqrt{1 + \frac{1}{n} + \frac{\left(x - \overline{x}\right)^2}{S_{xx}}}$$

where P(*T* ≤ *a*)= (1+*p*)/2 and *T* ~ *t*(*n* - 2).

The interval is called a 100*p*% prediction interval instead of a confidence interval, since *Y* is not a parameter but a random variable.

STAT 230/231 Example

A 95% prediction interval for the STAT 231 mark for a randomly chosen student who obtained a mark of 75 in STAT 230 is

$$\hat{\alpha} + \hat{\beta}x \pm a s_e \sqrt{1 + \frac{1}{n} + \frac{\left(x - \bar{x}\right)^2}{S_{XX}}}$$

$$= -4.0667 + 0.9944(75)$$

$$\pm 2.0484(9.4630)\sqrt{1 + \frac{1}{30} + \frac{\left(75 - 76.7333\right)^2}{5135.8667}}$$

$$= 70.51 \pm 19.7100$$

or [50.8, 90.2]

STAT 230/231 Example

A 95% prediction interval for the STAT 231 mark for a randomly chosen student who obtained a mark of 60 in STAT 230 is

$$\hat{\alpha} + \hat{\beta}x \pm a s_e \sqrt{1 + \frac{1}{n} + \frac{\left(x - \bar{x}\right)^2}{S_{XX}}}$$

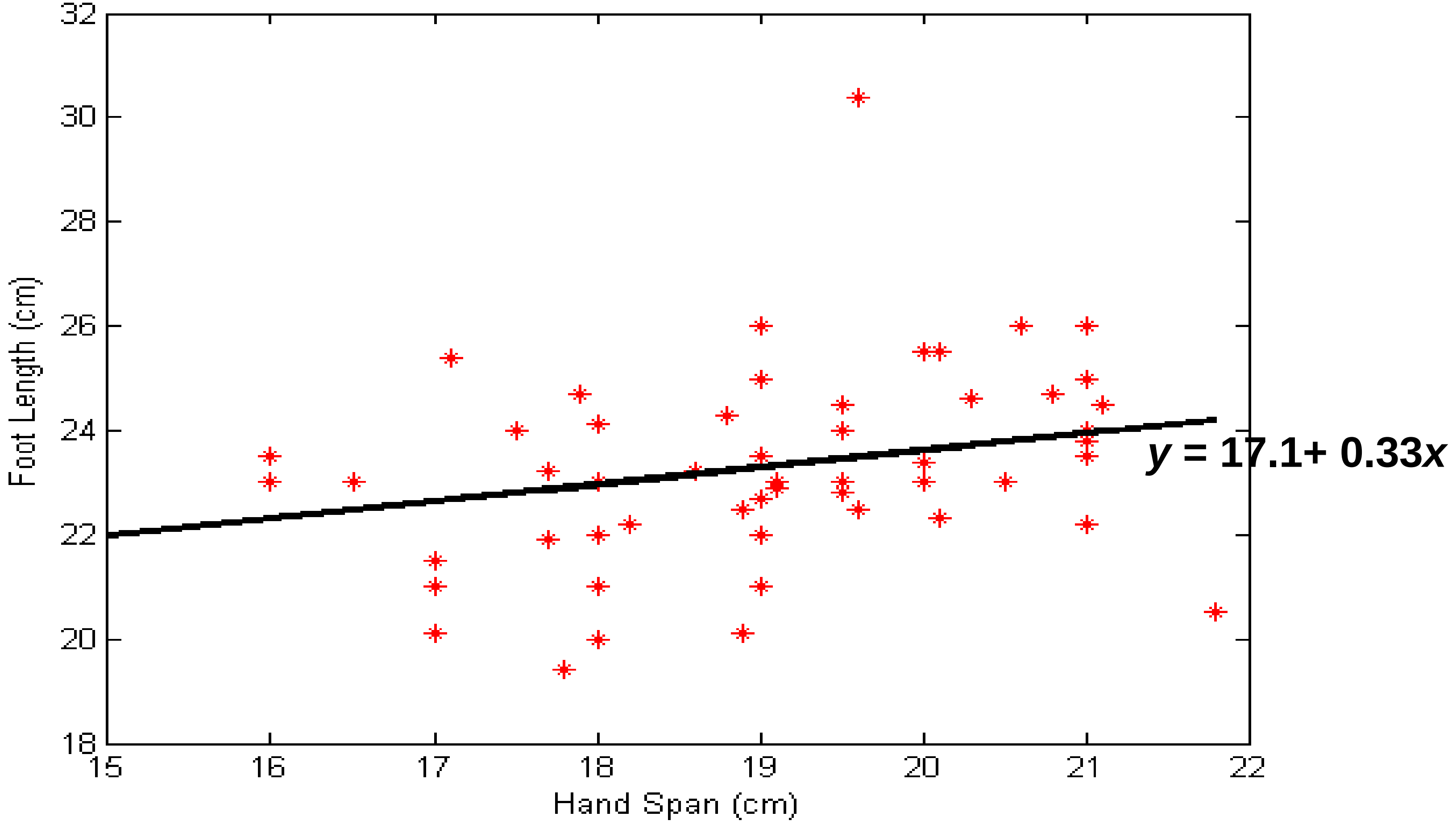
$$= -4.0667 + 0.9944(60)$$

$$\pm 2.0484(9.4630)\sqrt{1 + \frac{1}{30} + \frac{\left(60 - 76.7333\right)^2}{5135.8667}}$$

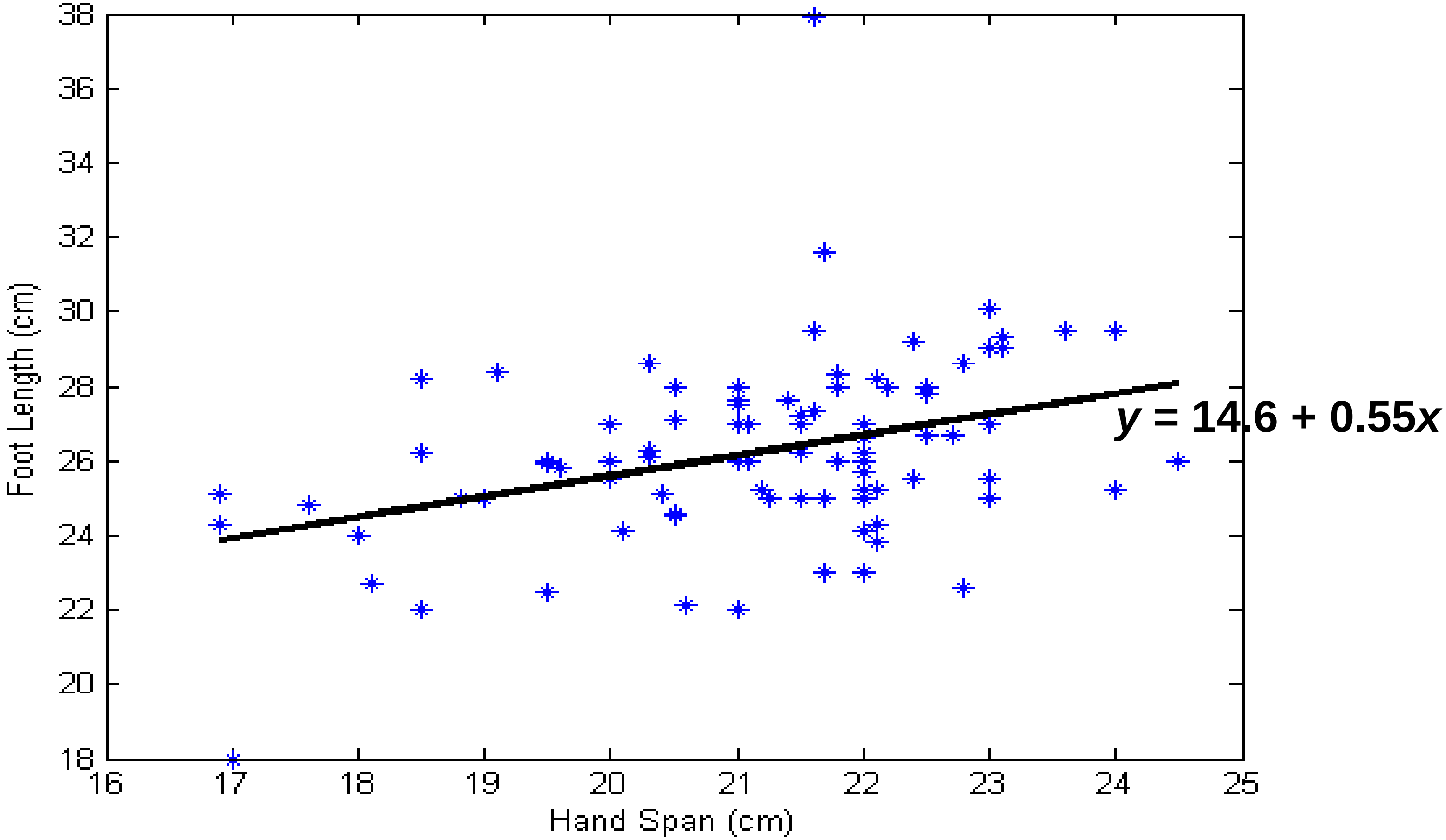
$$= 55.59 \pm 20.22$$

or [35.4, 75.8]

Foot Length versus Hand Span - Females



Foot Length versus Hand Span - Males



Foot Length versus Hand Span – Males (Blue), Females (Red)