Ch 15 Inference for Regression — Activity intro

Activity 15 Ideal Proportions (pg. 888)
"The length of a man's outspread arms is equal to his height."

1. x = distance between fingertips of outstretched arms
   y = height (in cm)
2. create scatterplot
   strong, moderate, weak association?
3. Find the linear regression line, r and r^2.
   Interpret meaning in context.
   r = correlation; $[-1, 1]$ strong, mod, weak; +/−
   r^2 = coefficient of determination; % of variation in y that can be explained by the least-squares line of y on x

   $\hat{y} = a + bx$
4. Construct a residual plot

Calculator

Least-Squares line: Data ⇒ L_1 & L_2
$\boxed{\text{STAT}}$ $\boxed{\text{CALC}}$ LinReg a+bx (L_1, L_2, Y_1)
Y_1 ⇒ $\boxed{\text{Vars}}$ $\boxed{\text{Y-Vars}}$ $\boxed{\text{Function}}$ Y_1
r^2/r ⇒ $\boxed{\text{Catalog}}$ Diagnostic On
$\boxed{\text{Graph}}$ ZOOM STAT

Residual Plot: Clear $y=$
L_3: $Y_1(L_1)$ L_4: $L_2 - L_3$ ⇒ Residuals in L_4
Stat Plot x: L_2 Y: L_4
Zoom Stat

Least Squares Regression (LSR) line can be used
to predict y-values for a given x

parameter		estimate
μ	mean	$\bar{x}$
σ	stand. dev.	s
p	proportion	$\hat{p}$
$\mu_y = \alpha + \beta x$	LSR	$\hat{y} = a + bx$

Example 15.1 Enter data into calculator Crying L_1, IQ L_2
Graph Scatter plot / LSR

$\hat{y} = 91.268 + 1.493x$ Read through example independently

True Regression Line = Line of Means or Line of Averages
$\boxed{\mu_y = \alpha + \beta x}$
mean response of y moves along the line as the explanatory variable x takes different values

Graphic (pg 892) - All of the curves have the same σ, so the variability of y is the same for all values of x.

Conditions for Regression Inference LINeSS
- L - linear; Scatterplot is roughly linear
- I - independent; no item is measured twice
- N - normal; boxplot of residuals is ~N
- S - SRS
- S - Standard Deviation; σ of y is the same for all values of x.; No fanning of the residual plot

15 cont'd

Example 15.2 Baby crying and IQ

$\hat{y} = 91.268 + 1.493x$

(Diff from text) (A) Two babies had 9 crying peaks.

Find their predicted IQ's and residuals

Predicted IQ: $\hat{y} = 91.27 + 1.493(9)$

$= 104.707$

Residuals: $103 - 104.707 = -1.707$

actual lies 1.707 points below LSR

$119 - 104.707 = 14.293$

actual lies 14.293 above LSR

(B) Determine all residuals

Calculator: Crying - L_1 IQ - L_2

$L_3 = 91.268 + 1.493(L_1)$

$L_4 = L_2 - L_3$

Stemplot pg. 844

$\boxed{\text{1-VAR Stats}}$ L_4 $\Sigma = -.057$ Sum is ~0

(Equal distribution +/-)

$\boxed{\text{Residual Plot}}$ StatPlot: On: X: L_2 Y: L_4

Zoom Stat

Standard Error of the LSR

$$s = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}} = \sqrt{\frac{\Sigma \text{residuals}^2}{n-2}}$$

* used in computation for CI of slope

Calculator $S = \sqrt{\frac{\Sigma \text{residuals}^2}{n-2}} \Rightarrow S^2 = \frac{\Sigma \text{residuals}^2}{n-2}$

L4 has all residuals from Ex 15.2

$\boxed{\text{1 VAR Stat}}$ L4 $\Rightarrow$ $\Sigma x^2 = 11023.389$

$n = 38$

$S^2 = \frac{11023.389}{38-2} = 306.20525$ $S = \sqrt{306.20525}$

$= 17.499$

$\boxed{S = 17.499}$

Confidence Interval for Regression Slope $b \pm t^* SE_b$

b = slope estimate $\hat{y} = a + bx$

t^* = critical value @ CI% for $df = n-2$

$\text{invT}(\frac{1+CI}{2}, df)$ or Table C

$SE_b = S_b = \frac{\sqrt{\frac{\Sigma (y-\hat{y})^2}{n-2}}}{\sqrt{\Sigma (x-\bar{x})^2}}$ * on formula sheet

* use list functions to complete the formula

Review

* Figure 15.5 Minitab output
* Table C t^* $df = n-2 = 38-2 = 36$ use $df = 30$

95% CI $\Rightarrow t^* = 2.042$

* Calculator $\text{invT}(.975, 36) = 2.20281$

HW 15.8

* Discuss 2 on scatterplot
* minitab output

15 cont'd

Steps for Constructing a Confidence Interval for the Regression Slope.

1. Procedure: t - confidence interval for the slope of the true regression line relating _____ and _____ for all ______.

2. Check Conditions LINeSS

3. Computation

$b_1 \pm t^* SE_{b_1}$ on formula sheet

$$SE_{b_1} = S_{b_1} = \frac{s}{\sqrt{\Sigma(x-\bar{x})^2}} = \frac{\sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}}}{\sqrt{\Sigma(x-\bar{x})^2}}$$

4. Conclusion:

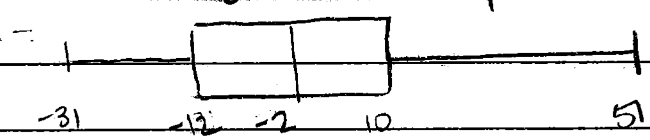
Based on this sample, I am ___ % confident that the slope of the true LSR relating ____ and ____ for all _____ is between ____ and ____.

Example

Construct a 95% confidence interval for the slope of the true regression lines of crying babies and IQ.

① We will use a t-confidence interval for the slope of the true regression line relating crying events and IQ for all infants.

② L - The scatterplot is roughly linear
I - We assume each infant was only tested once and independence is satisfied.
N -



The boxplot of residuals although skewed slightly to the right we will proceed as ~N

SRS is stated
The standard deviation of y is approximately the same for all x-values. The scatterplot of the residuals is ~N with no fanning.

③ $b_1 \pm t^* SE_{b_1}$

$b = 1.493 \quad df = 38 - 2 = 36 \quad t^* = 2.028$

$SE_{b_1} = \dfrac{\sqrt{\dfrac{\Sigma(y-\hat{y})^2}{n-2}}}{\sqrt{\Sigma(x-\bar{x})^2}}$

$\bar{x} = 17.395$

$1.493 \pm 2.028(.487)$

$(2.481, .505)$

$= \dfrac{\sqrt{\dfrac{11023.3889}{38-2}}}{\sqrt{1291.079}} = \boxed{.487}$

④ Based on the sample, I am 95% confident that the slope of the true LSR relating infant crying peaks and IQ at age 3 for all infants snapped with a rubber band then monitored for crying for 20 minutes is between 2.481 and .505.

Ch. 15 cont'd

* A linear relationship may not exist between two variables, like phone number and SSN. However, it is unlikely $b_1 = 0$ and each sample will likely produce a different b_1. Estimated slope is insignificant

t-Test for Slope
Determines if a nonzero correlation represents a true linear relationship with a nonzero slope and so knowing the value of x is helpful in predicting y.

Significance test for Regression Slope
"Is that trend real, or could that happen by chance?"
pg. 904 Graphic
"How far is b_1 from 0 (or some other H_0) in terms of standard error?"

Steps in t-Test for Slope
(1) Hypothesis: $H_0: \beta_1 = 0$ $H_a: \beta \gtrless\neq 0$
(2) Procedure: t-test for the slope of the true regression line (β_1) relating ____ & ____ for all ______
(3) Conditions: LINeSS
(4) Computation: $t = \frac{b_1 - \beta_1}{S_{b_1}}$ df = n−2   p-value = ____
(5) Interpret in context

Example! Temperature and cricket chirps (15 part II slide)

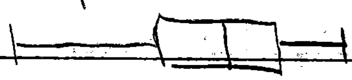
① $H_0: \beta_0 = 0$ $H_a: \beta_a \neq 0$

② t-test for the slope of the true regression line (β_1) relating the number of cricket chirps and the temperature for all one second intervals.

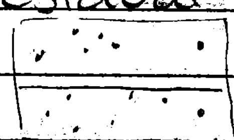
③ L - The scatterplot appears linear

Independent / SRS - We will consider this a random, independent sample of 1-sec. intervals

N_e - Box plot of residuals is $\sim N$



S - The residual plot shows no fanning



④ $t = \dfrac{b_1 - \beta}{S_{b_1}}$ $\quad b_1 = 3.291$ $\quad \beta = 0$ $\quad S_{b_1} = \dfrac{\sqrt{\frac{190.547}{13}}}{\sqrt{40.557}} = .601$

$t = \dfrac{3.291 - 0}{.601} = 5.475$

$df = 15 - 2 = 13$ $\quad$ p-value $= 2P(t > 5.475) \approx 0$

⑤ Since the p-value $0 < \alpha = .05$ we reject H_0. There is significant evidence to conclude temperature can be predicted by the number of cricket chirps.

(minitab output slides 15II and pg. 905)

$\boxed{\text{Ch. 15}}$ List Procedure for Regression Inference

L_1	L_2	L_3	L_4	L_5
X	Y	$Y_1(L_1)$	$L_2 - L_3$	$L_1 - \bar{x}$

LinReg (a+bx)
L_1, L_2
Store Eq: Y_1

Residuals
1 VAR Stat L_4
$\Sigma x^2 = \Sigma(y-\hat{y})^2$

$\bar{x} \Rightarrow$ 1 var stat L_1
1 VAR Stat L_5
$\Sigma x^2 = \Sigma(x-\bar{x})^2$

$$SE_{b_1} = S_{b_1} = \frac{\sqrt{\dfrac{\Sigma(y-\hat{y})^2}{n-2}}}{\sqrt{\Sigma(x-\bar{x})^2}}$$

$y - \hat{y} \Rightarrow$ residuals
$df = n-2$

<u>Box Plot of Residuals</u> - tests for Normality
StatPlot $\boxed{\text{on}}$
L_4

<u>Residual Plot</u> - test for consistent σ
StatPlot $\boxed{\text{on}}$ — no fanning
L_2, L_4