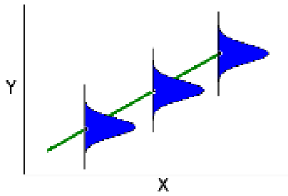
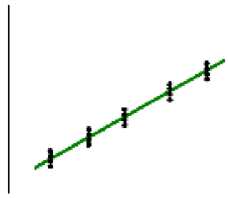
14.1 Inference About The Model

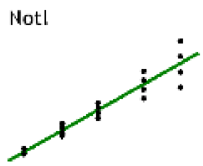
page 783

CONDITIONS FOR REGRESSION INFERENCE

Linear relationship (linear scatterplot?)
Independent observations (how's data produced?)
Normally distributed y-values at each x (Normal residuals?)
Equal variance in y-values at each x (residuals uniformly spread?)
And
Random (how's data produced?)

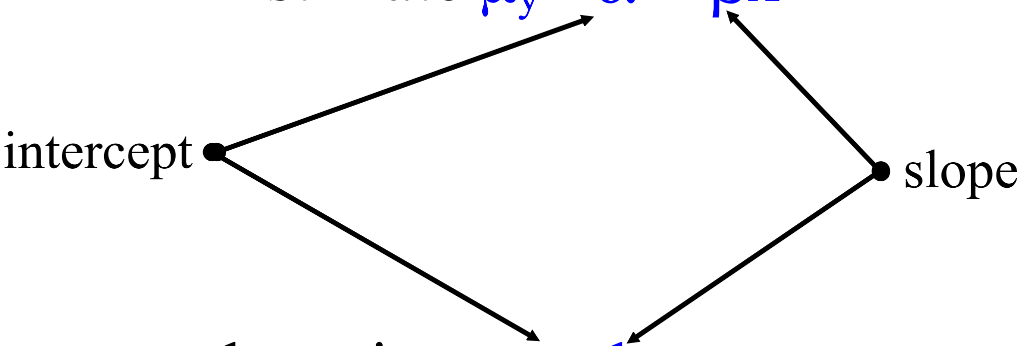






page 784

Estimate $\mu_y = \alpha + \beta x$



by using $y=a+bx$.

STANDARD ERROR ABOUT THE LEAST SQUARES LINE

page 784

Estimate unknown σ using s,
the standard error about the line:

$$s = \sqrt{\frac{\sum (residual)^2}{n-2}} = \sqrt{\frac{\sum (y-\hat{y})^2}{n-2}}$$

This says that if you would find slopes for many random samples, the sample slopes will be about s from the population slope, on average.

page 789

CI for the slope β of the regression line:
$(b - t^*SE_b\ ,\ b + t^*SE_b)$

The standard error of the least squares slope b is:

$$SE_b = \frac{s}{\sqrt{(x-\bar{x})^2}} = \sqrt{\frac{\sum(y-\hat{y})^2}{(n-2)\sum(x-\bar{x})^2}}$$

← SE_b is given in computer output

Note: t* uses n-2 degrees of freedom.

SIGNIFICANCE TESTS FOR REGRESSION SLOPE

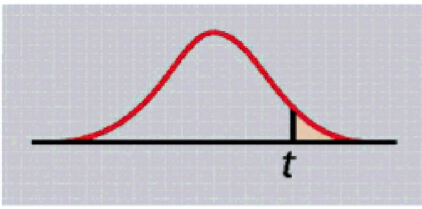
To test the hypothesis H_0: $\beta=0$, compute the t statistic

$$t = \frac{b-\beta}{SE_b}$$

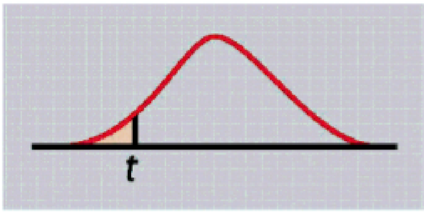
← at least 2 of these 3 parts are always on computer output

the P-value for a test of H_0 against

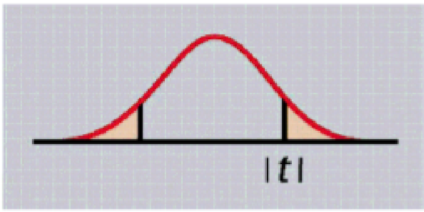
H_a: $\beta > 0$ is $P(T \geq t)$



H_a: $\beta < 0$ is $P(T \leq t)$



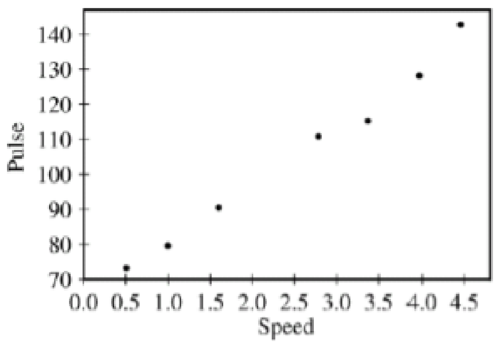
H_a: $\beta \neq 0$ is $2P(T \geq |t|)$



This also tests H_0: correlation is 0 in the population.

AP Statistics 2005 (B) #5

John believes that as he increases his walking speed, his pulse rate will increase. He wants to model this relationship. John records his pulse rate, in beats per minute (bpm), while walking at each of seven different speeds, in miles per hour (mph). A scatterplot and regression output are shown below.



Regression Analysis: Pulse Versus Speed

Predictor	Coef	SE Coef	T	P
Constant	63.457	2.387	26.58	0.000
Speed	16.2809	0.8192	19.88	0.000

S = 3.087   R-Sq = 98.7%   R-Sq (adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3763.2	3763.2	396.13	0.000
Residual	5	47.6	9.5		
Total	6	3810.9			

(a) Using the regression output, write the equation of the fitted regression line.

Predicted Pulse = 63.457 + 16.2809 (Speed)

Note: If the student uses *X* and *Y*, then both variables must be identified.

Part (a) is scored as essentially correct (E) or incorrect (I).

-

Regression Analysis: Pulse Versus Speed

Predictor	Coef	SE Coef	T	P
Constant	63.457	2.387	26.58	0.000
Speed	16.2809	0.8192	19.88	0.000

S = 3.087 R-Sq = 98.7% R-Sq (adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3763.2	3763.2	396.13	0.000
Residual	5	47.6	9.5		
Total	6	3810.9			

(b) Do your estimates of the slope and intercept parameters have meaningful interpretations in the context of this question? If so, provide interpretations in this context. If not, explain why not.

The intercept (63.457 bpm) provides an estimate for John's mean resting pulse (walking at a speed of zero mph).

The slope (16.2809 bpm/mph) provides an estimate for the mean increase in John's heart rate as his speed is increased by one mile per hour.

Part (b): There are four steps to constructing correct interpretations:

Step 1: A correct mathematical interpretation of the reported slope (16.2809) as a rate of increase in heart rate as walking speed increases.

Step 2: A correct mathematical interpretation of the reported intercept as a pulse rate when walking speed is zero.

Step 3: Correct use of units of measurement, e.g., John's heart rate increases 16.2809 bpm as his speed is increased by one mile per hour.

Step 4: Interpretation of the reported values as estimates of the corresponding mean quantities.

Part (b) is essentially correct (E) if all four steps are correct.

Part (b) is partially correct (P) if two or three steps are correctly addressed. Step 2 is scored as incorrect, for example, if the student suggests that the intercept does not have a meaningful interpretation.

Part (b) is incorrect (I) if at most one step is correct.

Note: The student is only penalized once for switching the variables.

Regression Analysis: Pulse Versus Speed

Predictor	Coef	SE Coef	T	P
Constant	63.457	2.387	26.58	0.000
Speed	16.2809	0.8192	19.88	0.000

S = 3.087 R-Sq = 98.7% R-Sq (adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3763.2	3763.2	396.13	0.000
Residual	5	47.6	9.5		
Total	6	3810.9			

(c) John wants to provide a 98 percent confidence interval for the slope parameter in his final report. Compute the margin of error that John should use. Assume that conditions for inference are satisfied.

Table B *t* distribution critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

The margin of error for the confidence interval for the slope parameter is $t^*_{n-2} \times s_b$, where s_b is the standard error of the slope parameter. For a 98% confidence interval, the margin of error is $3.365 \times 0.8192 = 2.7566$ bpm.

Part (c) is essentially correct (E) if the standard error of the slope is identified and the correct critical value is used to calculate the margin of error.

Part (c) is partially correct (P) if the student:

- Computes the 98% confidence interval but does not identify the margin of error; OR
- Recognizes that the margin of error consists of the standard error of the coefficient and the critical value but uses an incorrect value for one of the two components or uses a *t*-value with 6 degrees of freedom and an incorrect standard error.

Part (c) is incorrect (I) if the student uses:

- The standard error of the coefficient as the margin of error; OR
- A critical value as the margin of error.

14.1 MLB ht wt data 2005.ftm

Population: 2005 MLB players

β= slope of line relating ht. and wt. of 2005 MLB players

H_0: $\beta = 0$

H_a: $\beta > 0$

height	weight
76	246
72	207
75	220
74	200
72	170
71	175
68	150
74	210
74	245
72	200

Table B *t* distribution critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

14.1 MLB ht wt data 2005.ftm