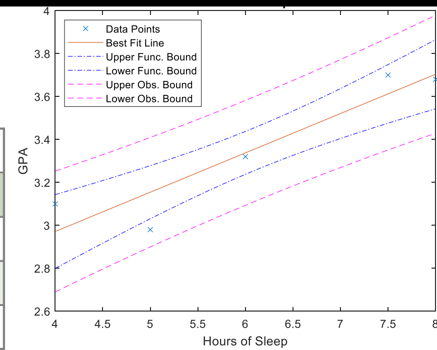
E80

Experimental Engineering

$$SSE = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} \left[y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i \right) \right]^2$$

SIGNIFICANCE LEVEL FOR TWO-TAILED TEST						
df	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610



Lecture 2E – Linear Regression

1

E80

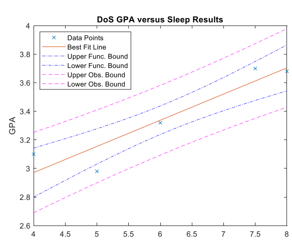
Experimental Engineering

Overview of Linear Regression

- Take data.
- Assume linear.
- Perform regression.
- Report slope and intercept.
- Plot data, line, uncertainties.

$$\hat{\beta}_1 = 0.183 \pm 0.067 (80\% \text{ conf.})$$

$$\hat{\beta}_0 = 2.238 \pm 0.418 (80\% \text{ conf.})$$



Partial list of techniques:

- Least Squares
- Least Absolute Deviation (LAD)
- Median-Median
- Least Median of Squares (LMS)

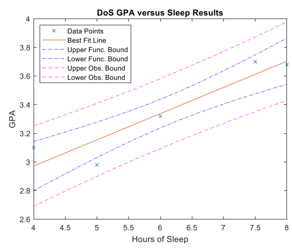
2

E80 Experimental Engineering — Overview of Linear Regression

- Take data.
- Assume linear.
- Perform regression.
- Report slope and intercept.
- Plot data, line, uncertainties.

$$\hat{\beta}_1 = 0.183 \pm 0.067 (80\% \text{ conf.})$$

$$\hat{\beta}_0 = 2.238 \pm 0.418 (80\% \text{ conf.})$$



Partial list of techniques:

- **Least Squares**
- Least Absolute Deviation (LAD)
- Median-Median
- Least Median of Squares (LMS)

3

E80 Experimental Engineering — Measurements and Residuals

The set of measurements

$$\{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Can we calculate the error?

$$\varepsilon_i = y_i - \left(\beta_0 + \beta_1 x_i\right)$$

The set of errors $\{\varepsilon_1, \varepsilon_2, \cdots \varepsilon_N\}$

Calculate the residual

$$e_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$$

The set of residuals

$$\{e_1, e_2, \cdots e_N\}$$

The Sum of Squared rEsiduals, SSE

$$\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{N} \left[y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right]^2$$

4

E80 Experimental Engineering

Best-Fit Slope & Intercept

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Root Mean Squared rEsidual ($RMSE$)

$$RMSE = S_e = \sqrt{\frac{SSE}{N-2}} = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N-2}}$$

Lose two degrees of freedom by calculating $\hat{\beta}_0, \hat{\beta}_1$

5

E80 Experimental Engineering

Standard Errors & C.I.'s

Standard Errors

$$S_{\beta_0} = S_e \sqrt{\frac{1}{N} + \frac{\bar{x}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}$$

$$S_{\beta_1} = S_e \sqrt{\frac{1}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}$$

Confidence Intervals

$$\lambda_{\beta_0} = t S_{\beta_0}$$

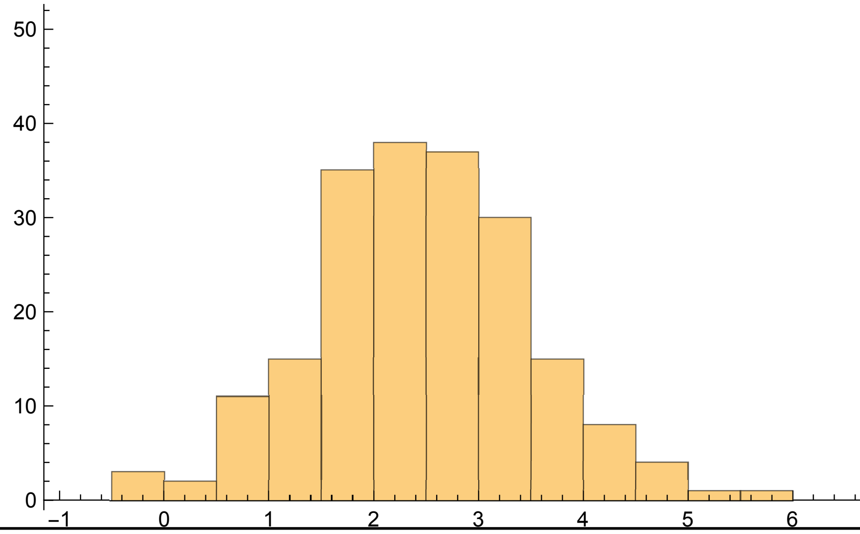
Remember: $df = N - 2$

$$\lambda_{\beta_1} = t S_{\beta_1}$$

6

E80
Experimental Engineering

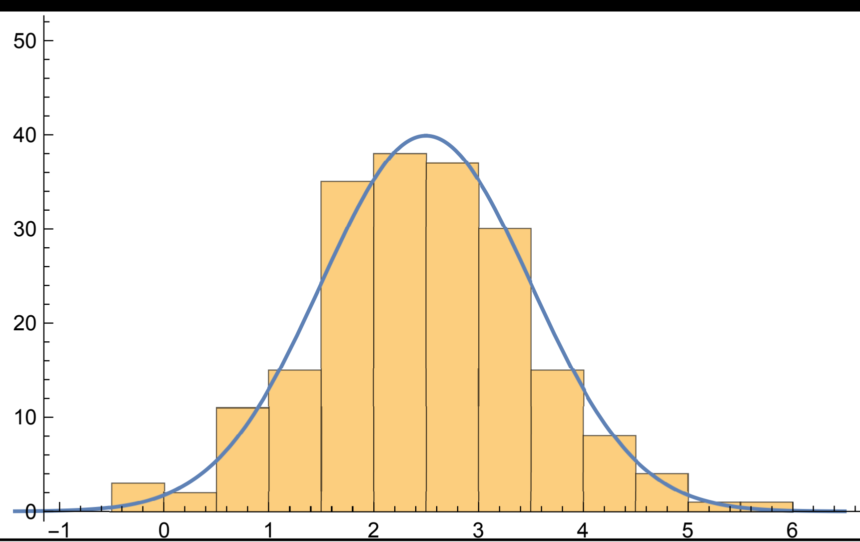
Histograms to Lines



7

E80
Experimental Engineering

Histograms to Lines

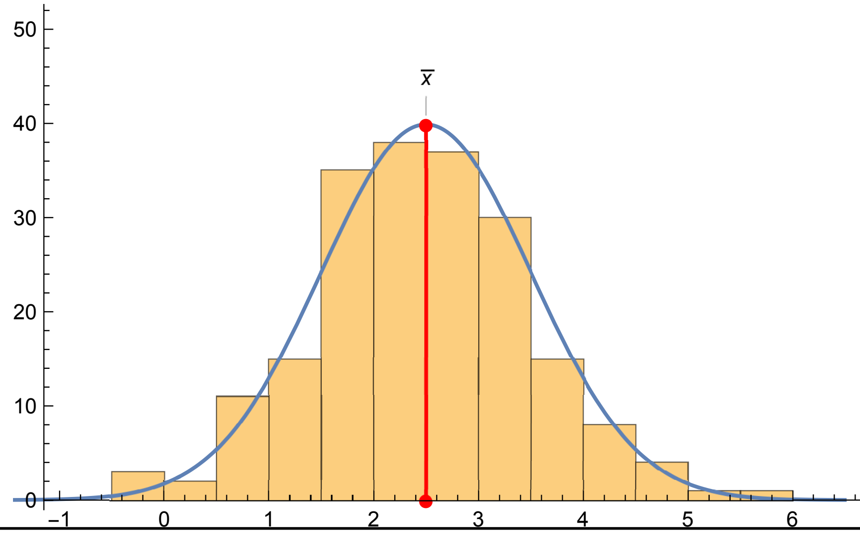


8

E80

Experimental Engineering

Histograms to Lines

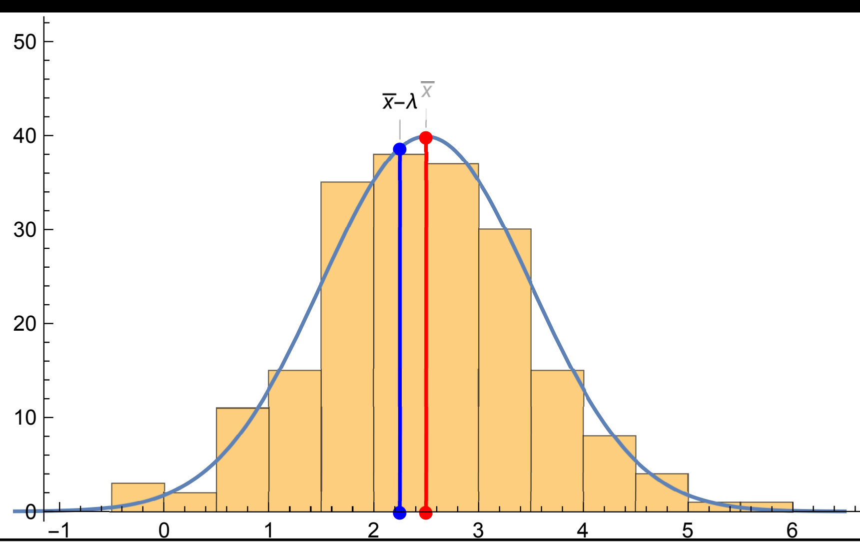


9

E80

Experimental Engineering

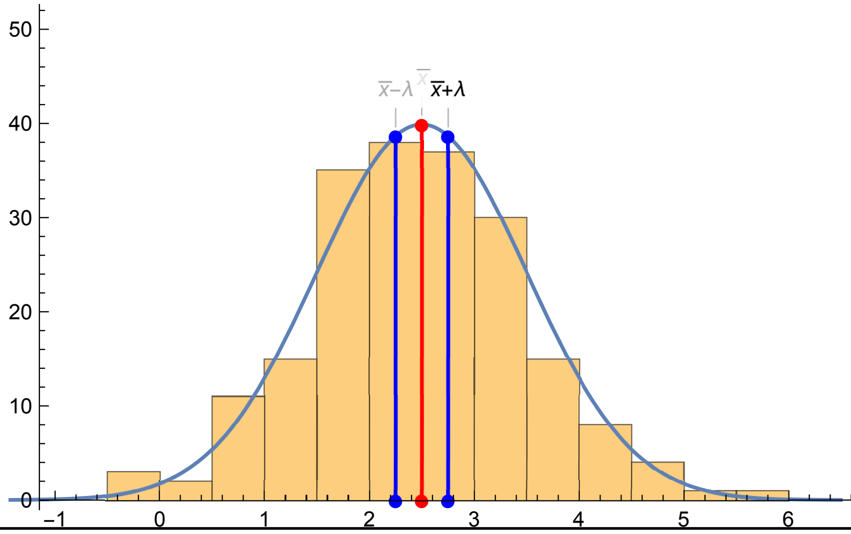
Histograms to Lines



10

E80
Experimental
Engineering

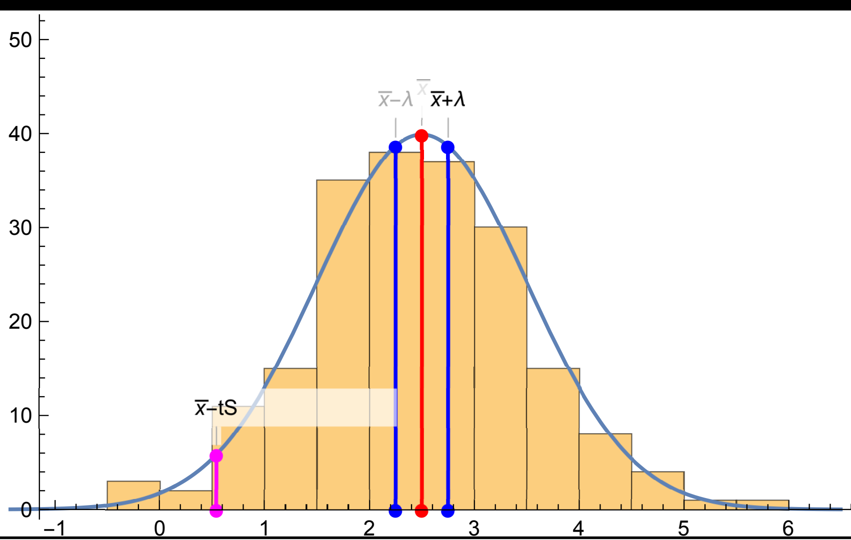
Histograms to Lines



11

E80
Experimental
Engineering

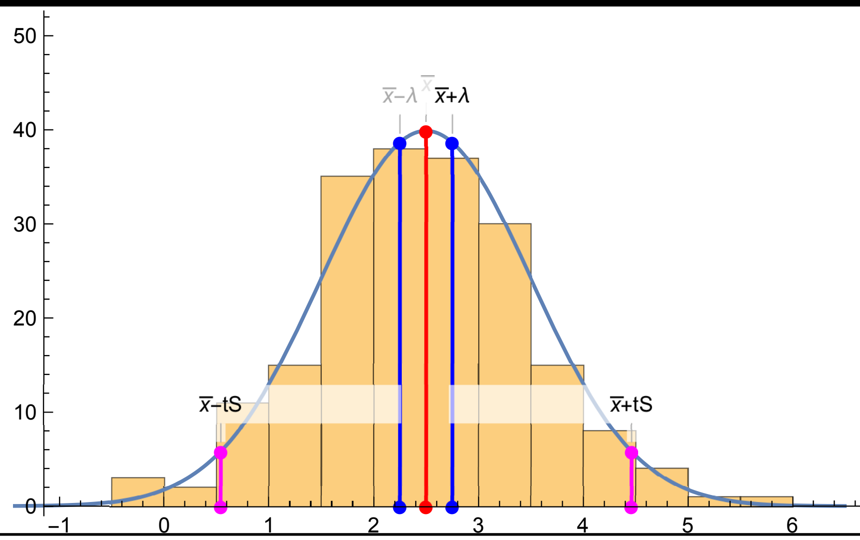
Histograms to Lines



12

E80
Experimental
Engineering

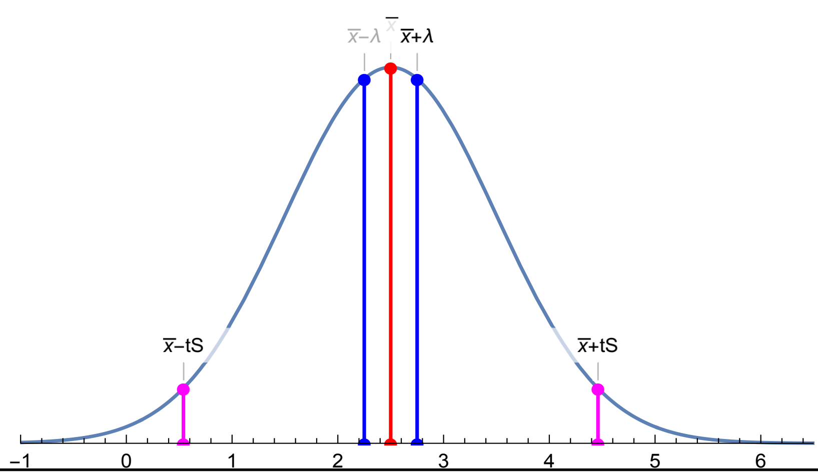
Histograms to Lines



13

E80
Experimental
Engineering

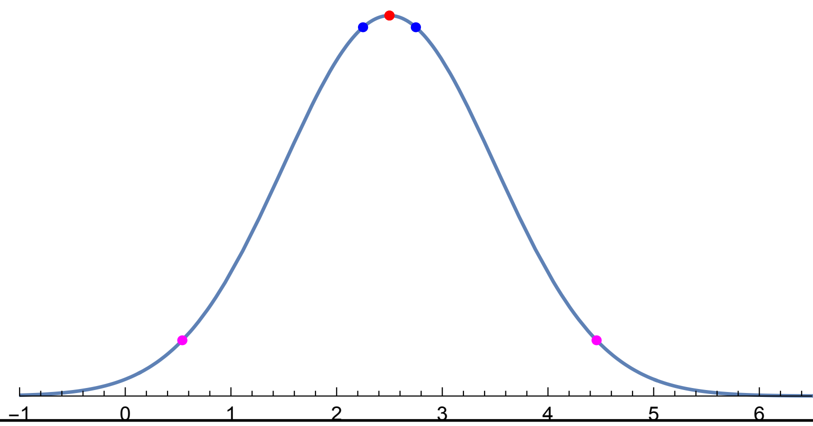
Histograms to Lines



14

E80
Experimental Engineering

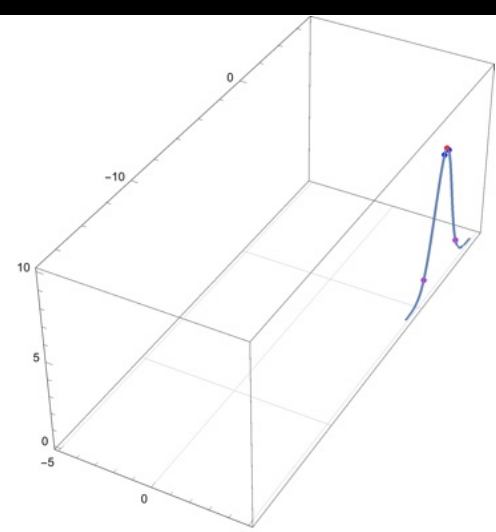
Histograms to Lines



15

E80
Experimental Engineering

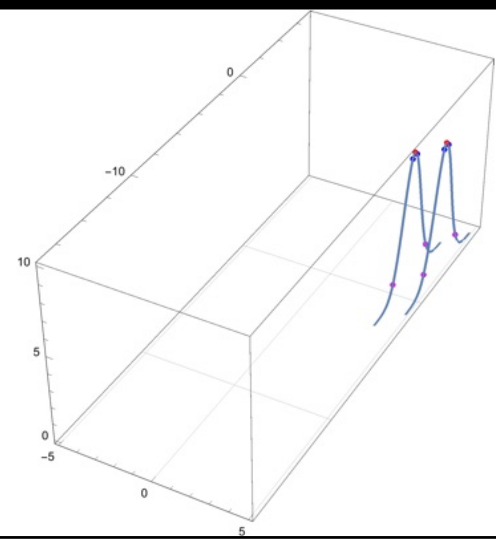
Histograms to Lines



16

E80
Experimental Engineering

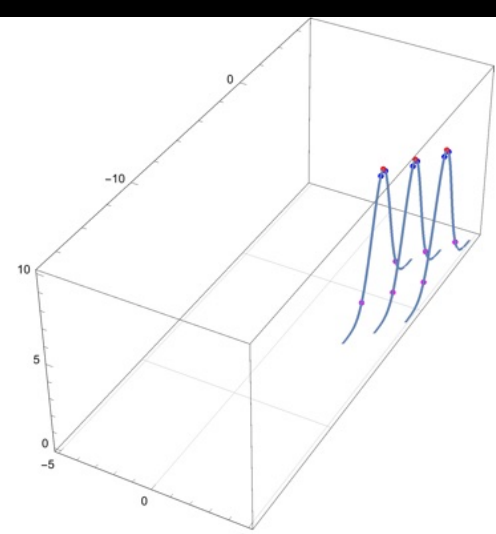
Histograms to Lines



17

E80
Experimental Engineering

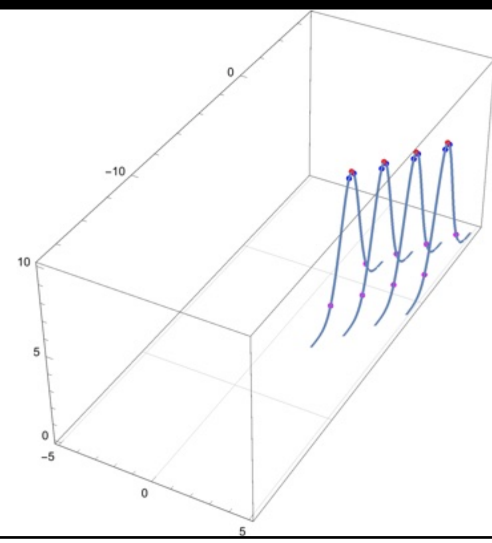
Histograms to Lines



18

E80
Experimental
Engineering

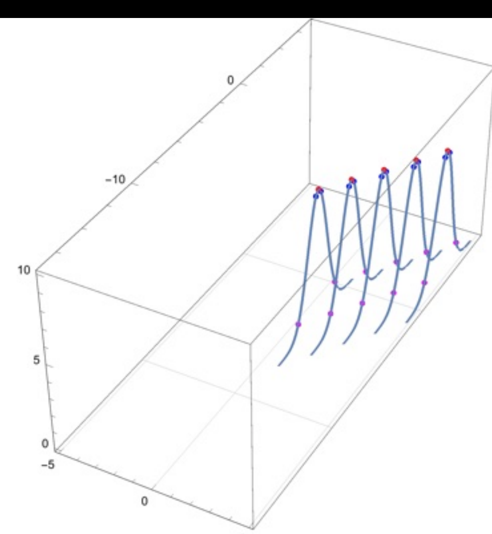
Histograms to Lines



19

E80
Experimental
Engineering

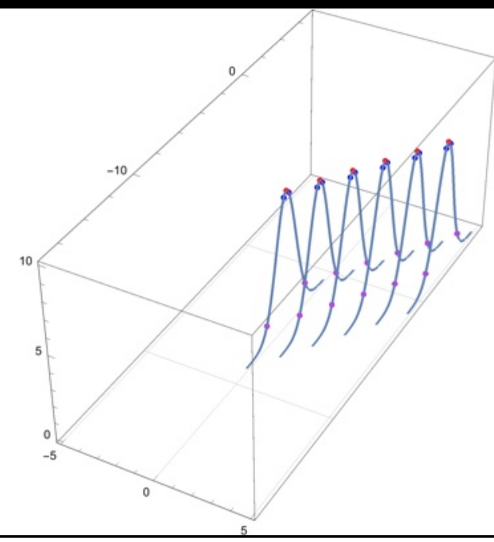
Histograms to Lines



20

E80
Experimental
Engineering

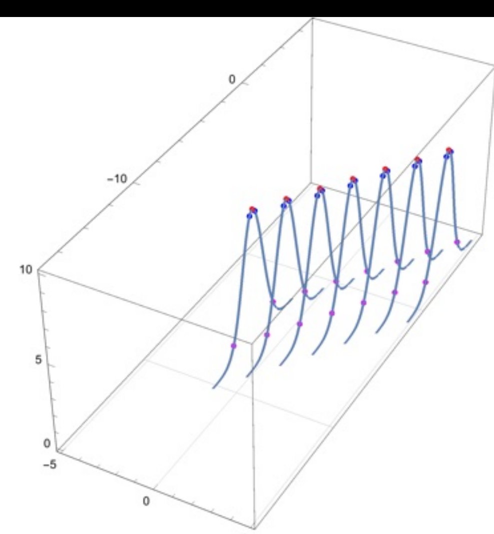
Histograms to Lines



21

E80
Experimental
Engineering

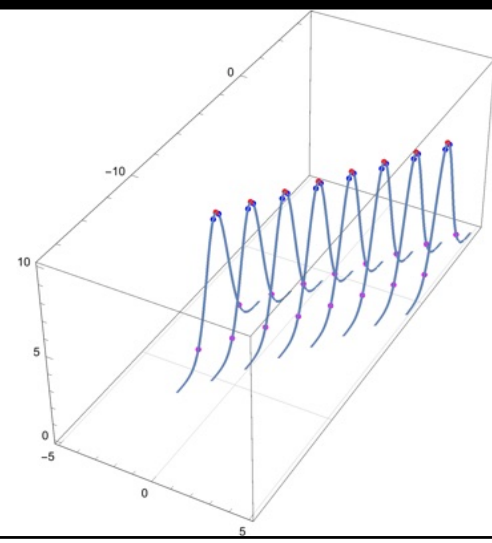
Histograms to Lines



22

E80
Experimental Engineering

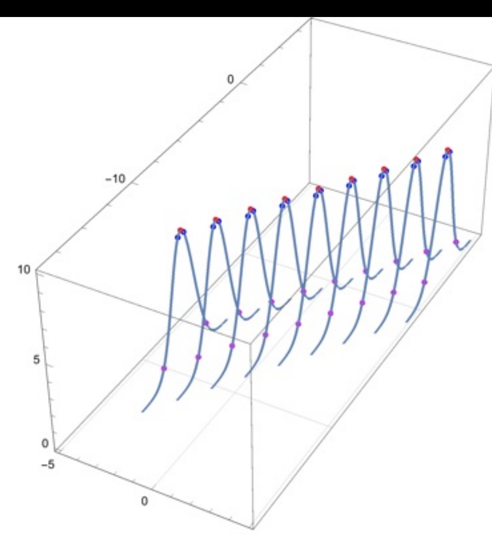
Histograms to Lines



23

E80
Experimental Engineering

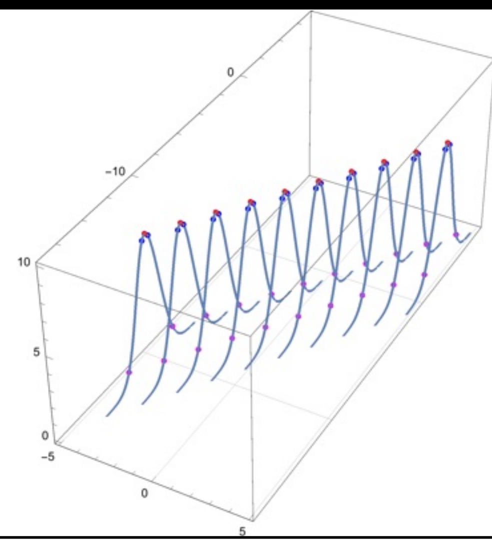
Histograms to Lines



24

E80
Experimental Engineering

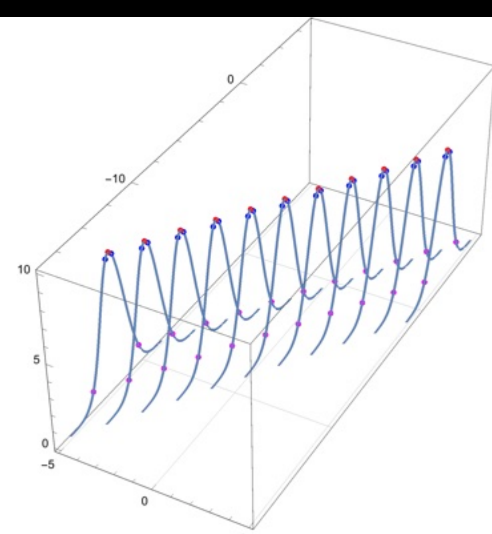
Histograms to Lines



25

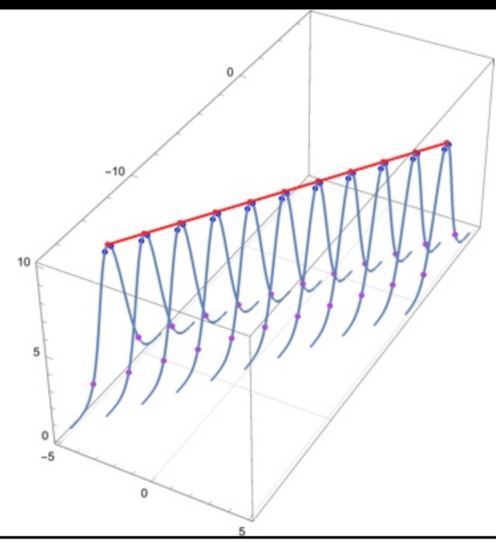
E80
Experimental Engineering

Histograms to Lines



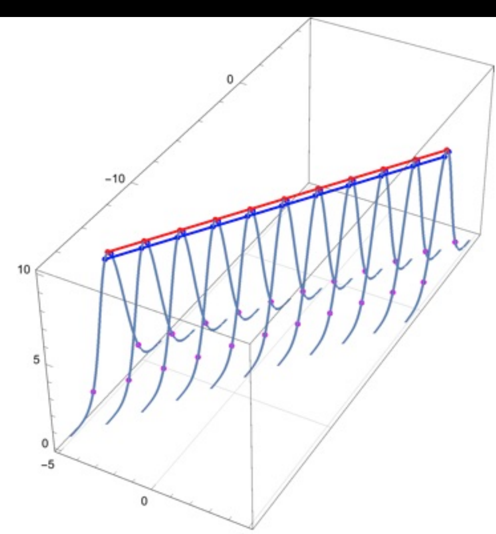
26

E80 Experimental Engineering — Histograms to Lines



27

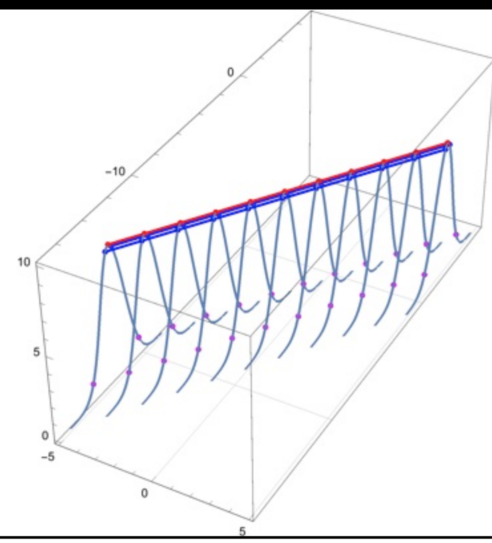
E80 Experimental Engineering — Histograms to Lines



28

E80 Experimental Engineering

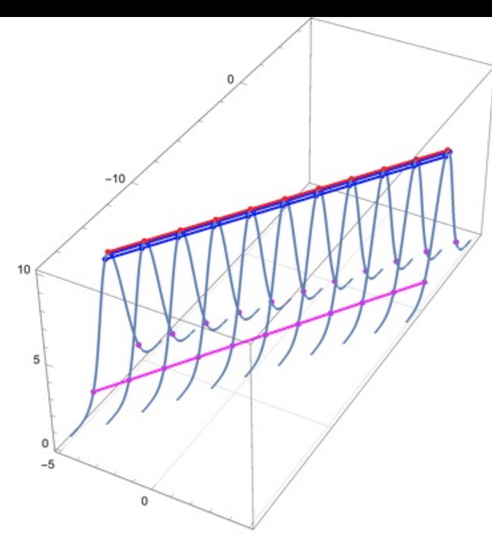
Histograms to Lines



29

E80 Experimental Engineering

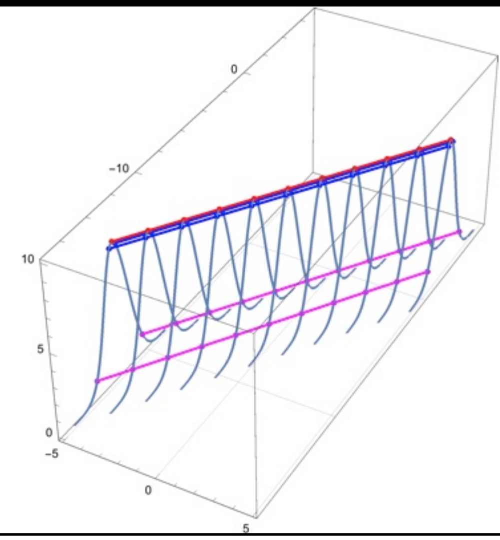
Histograms to Lines



30

E80 Experimental Engineering

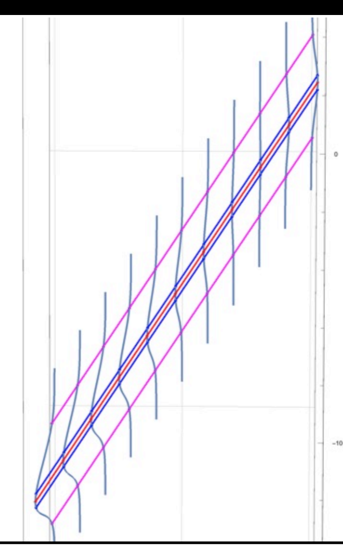
Histograms to Lines



31

E80 Experimental Engineering

Histograms to Lines



32

E80
Experimental Engineering

Confidence Bounds

Confidence bound on line or mean value $\hat{y}$ at x (functional bound)

$$S_{\hat{y}} = S_e \sqrt{\frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} \qquad \lambda_{\hat{y}} = tS_{\hat{y}}$$

$$y_{ucb\hat{y}} = \hat{\beta}_0 + \hat{\beta}_1 x + \lambda_{\hat{y}}$$

$$y_{lcb\hat{y}} = \hat{\beta}_0 + \hat{\beta}_1 x - \lambda_{\hat{y}}$$

Confidence bound on data (observational bound)

$$S_y = S_e \sqrt{1 + \frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} \qquad \lambda_y = tS_y$$

$$y_{ucby} = \hat{\beta}_0 + \hat{\beta}_1 x + \lambda_y$$

$$y_{lcby} = \hat{\beta}_0 + \hat{\beta}_1 x - \lambda_y$$

33

E80
Experimental Engineering

Example

	x_i (Hours of Sleep)	y_i (GPA)
	7.5	3.70
	4.0	3.10
	6.0	3.32
	5.0	2.98
	8.0	3.68

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = 0.183 \qquad S_{\beta_1} = S_e \sqrt{\frac{1}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} = 0.0407$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2.238 \qquad S_{\beta_0} = S_e \sqrt{\frac{1}{N} + \frac{\bar{x}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}} = 0.255$$

$$S_e = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N-2}} = 0.136 \qquad t(df = 3, 80\% \text{ conf.}) = 1.638$$

34

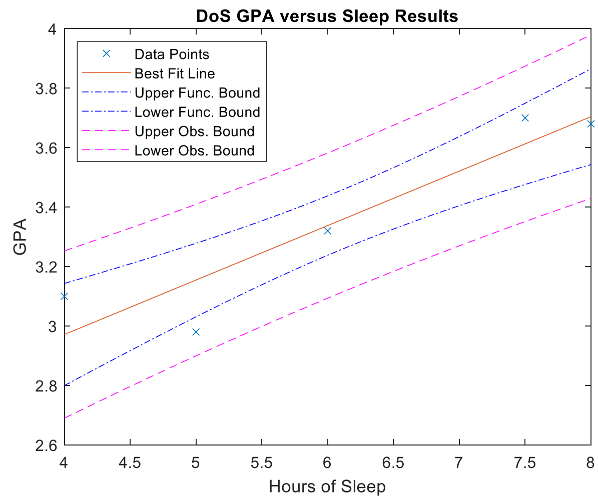
E80
Experimental Engineering

Example (cont.)

	x_i (Hours of Sleep)	y_i (GPA)
	7.5	3.70
	4.0	3.10
	6.0	3.32
	5.0	2.98
	8.0	3.68
mean	**6.1**	**3.36**

$$\hat{\beta}_1 = 0.183 \pm 0.067 (80\% \text{ conf.})$$

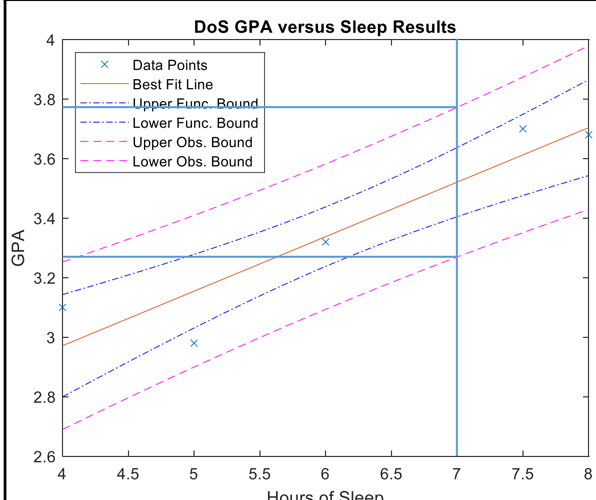
$$\hat{\beta}_0 = 2.238 \pm 0.418 (80\% \text{ conf.})$$



35

E80
Experimental Engineering

Example (cont.)



$$y = 2.238 + 0.183 \bullet 7 = 3.52$$

$$S_y = S_e \sqrt{1 + \frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}} = 0.154$$

$$\lambda_y = t S_y = 0.254$$

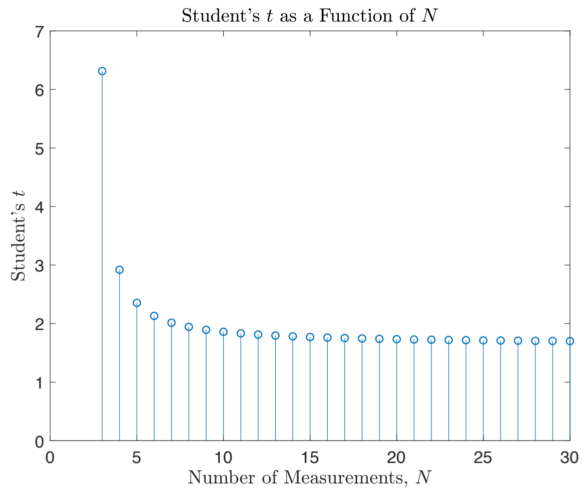
$$y(7) = 3.52 \pm 0.254 = \begin{bmatrix} 3.27 & 3.77 \end{bmatrix}$$

36

E80
Experimental
Engineering

What Does It Mean?

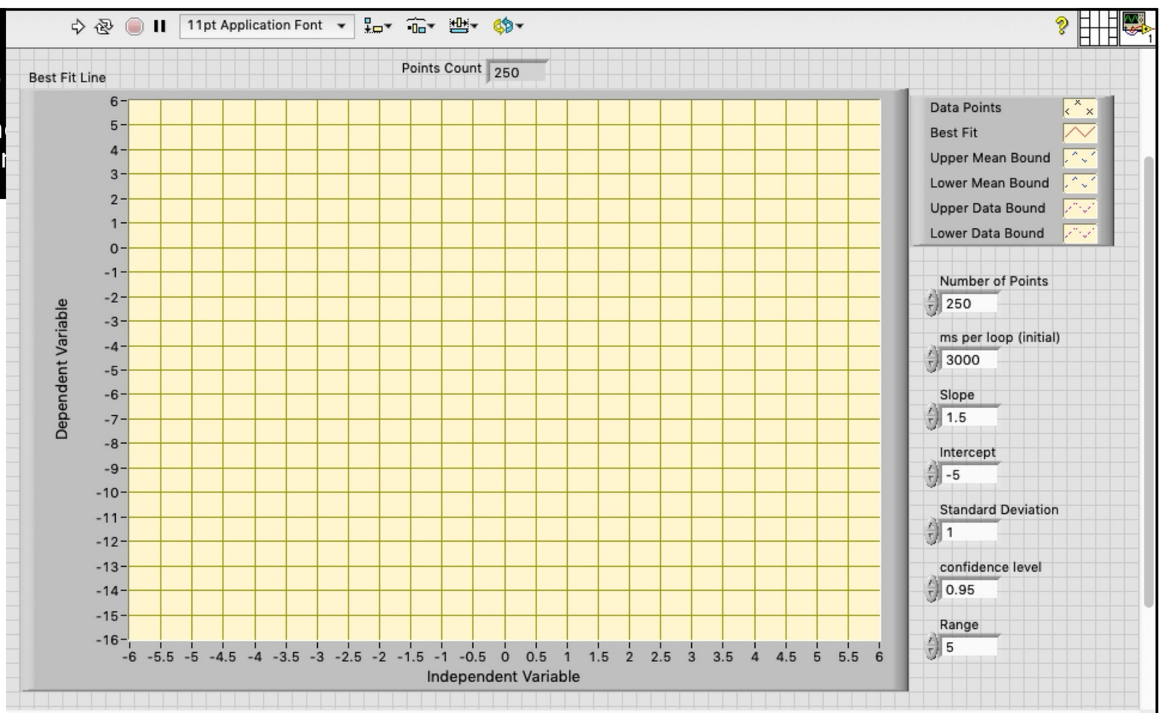
$$\lambda_{\hat{y}} = t S_{\hat{y}}$$



$$S_{\hat{y}} = S_e \sqrt{\frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}$$

37

E8
Experim
Engineer



38

E80

Experimental Engineering

Takaways

1. For a line, make at least four measurements.
2. DO NOT calculate these things by hand. Use a software package.
3. Report your results with the confidence interval and the confidence level, e.g., $\hat{\beta}_1 = 0.183 \pm 0.067 (80\% \text{ conf.})$ $\hat{\beta}_0 = 2.238 \pm 0.418 (80\% \text{ conf.})$
4. Plot data points, best-fit curve, function confidence bound, data confidence bound.

39