# Math 1710 Class 21

Correlation and Regression

Dr. Back

Oct. 16, 2009

# Dist of Sample vs. Sampling Dist



Underlying Probability Distribution

Theoretical

(leads to $\mu$, $\sigma$)

# Dist of Sample vs. Sampling Dist



Underlying Probability Distribution

Theoretical

(leads to $\mu$, $\sigma$)

Distribution of a Sample

Observational

(leads to $\bar{x}$, $s$)

# Dist of Sample vs. Sampling Dist



Sampling Distribution of $\bar{x}$

Theoretical

$(\mu,\ SD(\bar{x}) = \sigma/\sqrt{n})$

# Dist of Sample vs. Sampling Dist



Sampling Distribution of $\bar{x}$

Theoretical

$(\mu,\ SD(\bar{x}) = \sigma/\sqrt{n})$

Distribution of a few $\bar{x}$'s

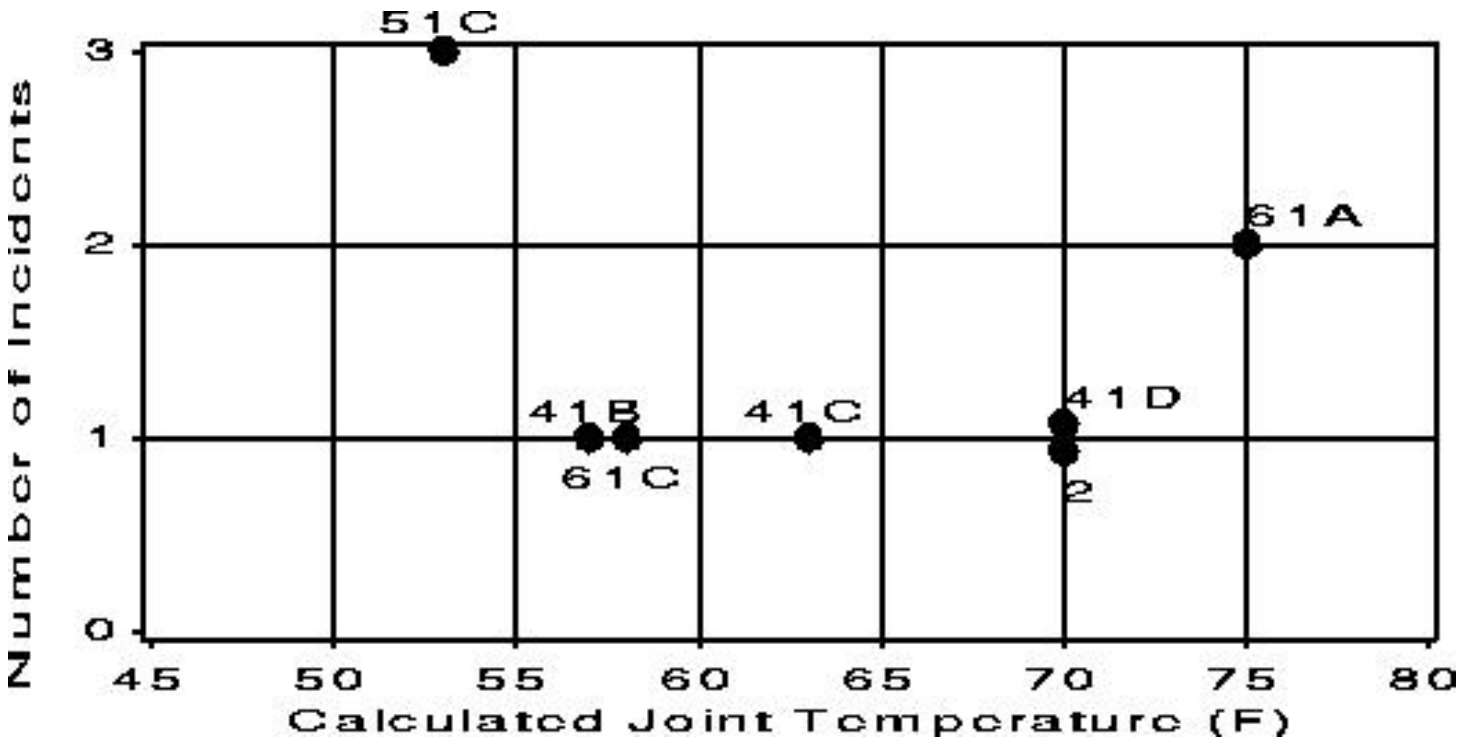
Observational version of sampling dist.

# Always Graph All the Data



Number of o-ring failures vs. temp. *if failures*

reviewed the night before

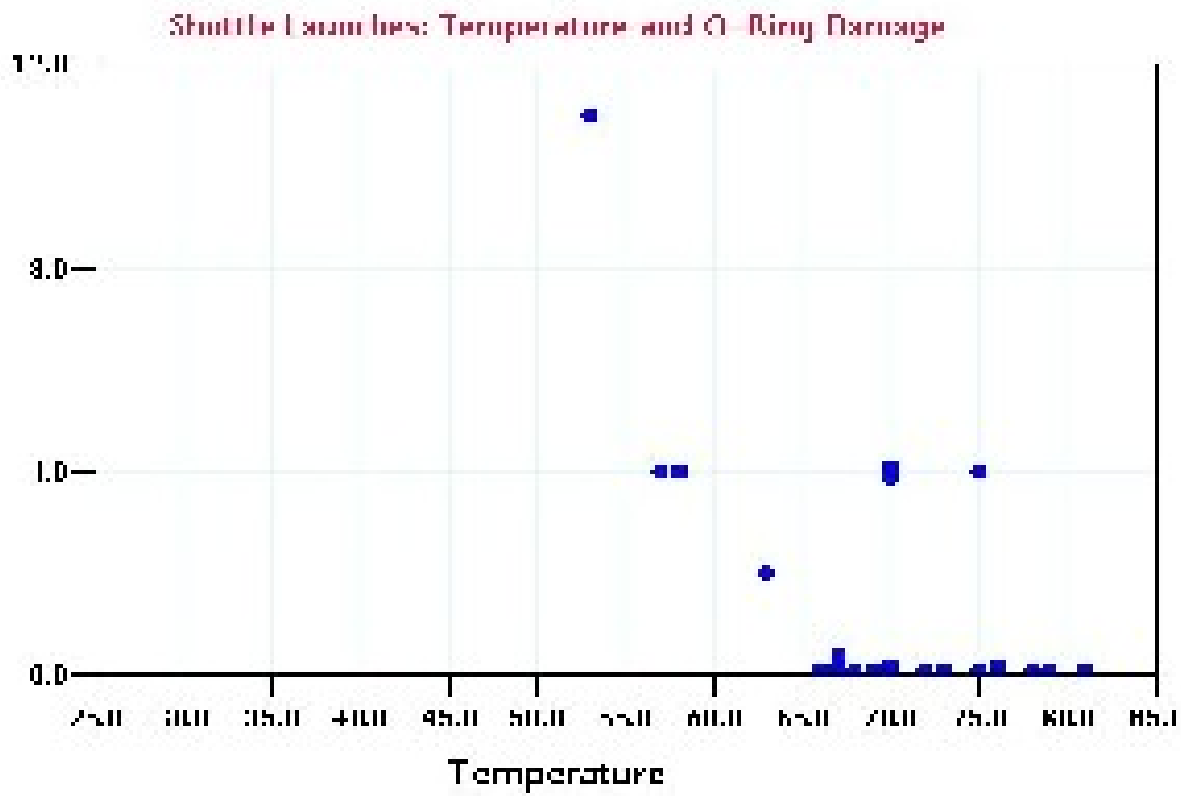


# Always Graph All the Data



Number of O ring failures vs. temp. *all missions*

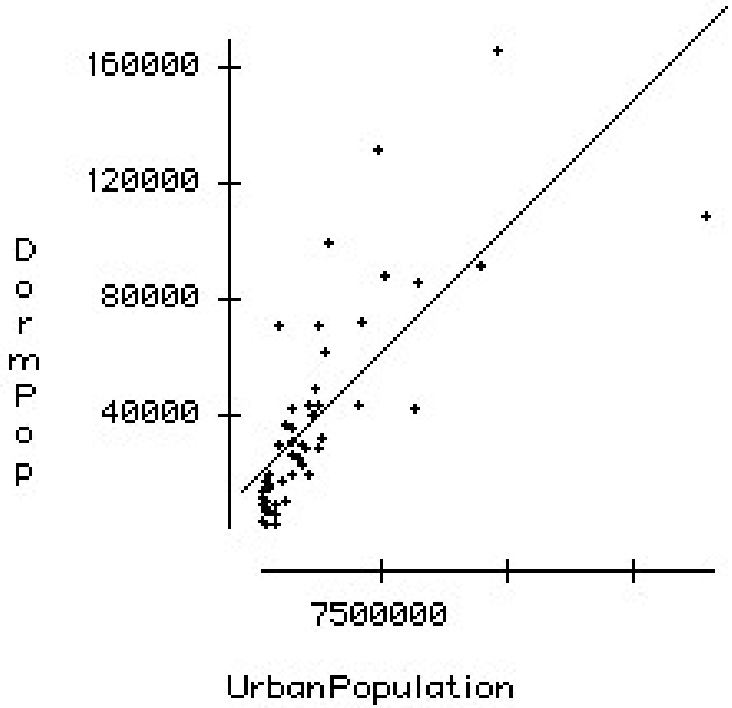
what should have been looked at



# Does an Urban Location Encourage Students to Live in Dorms?



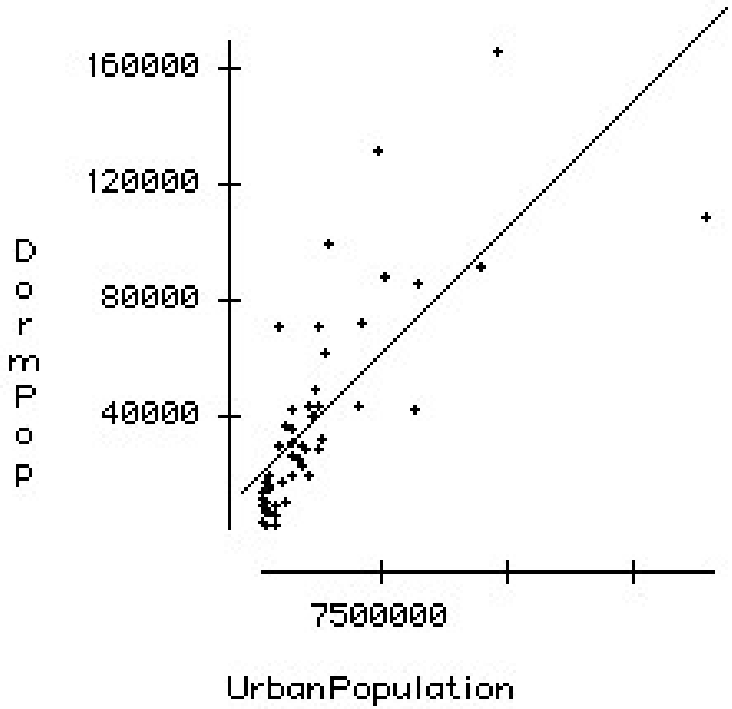
Dorm pop vs. Urban pop by State



# Does an Urban Location Encourage Students to Live in Dorms?



## Dorm pop vs. Urban pop by State




**Pearson Product-Moment Correlation**

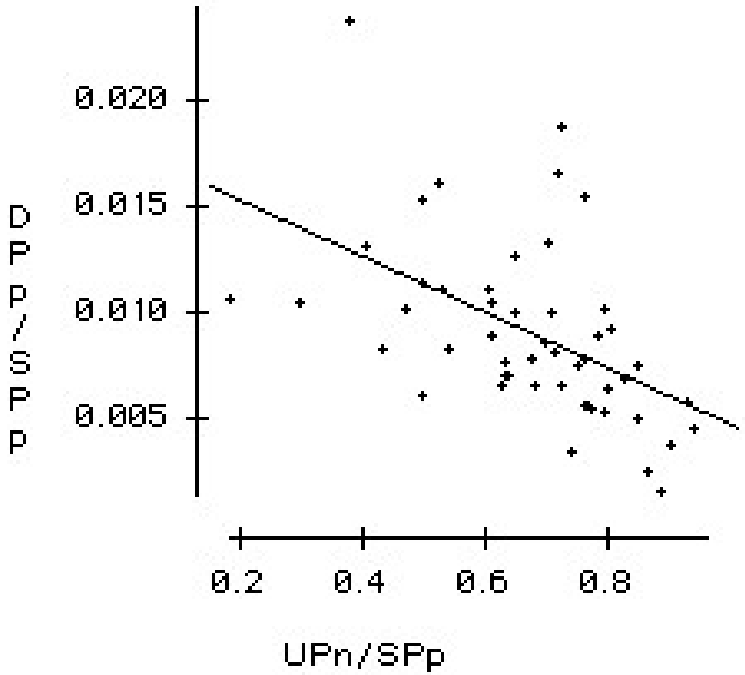
No Selector

| | DormP_ | UrbanP_ |
|---|---|---|
| **DormPop** | 1.000 | |
| **UrbanPopula_** | 0.752 | 1.000 |

# Does an Urban Location Encourage Students to Live in Dorms?



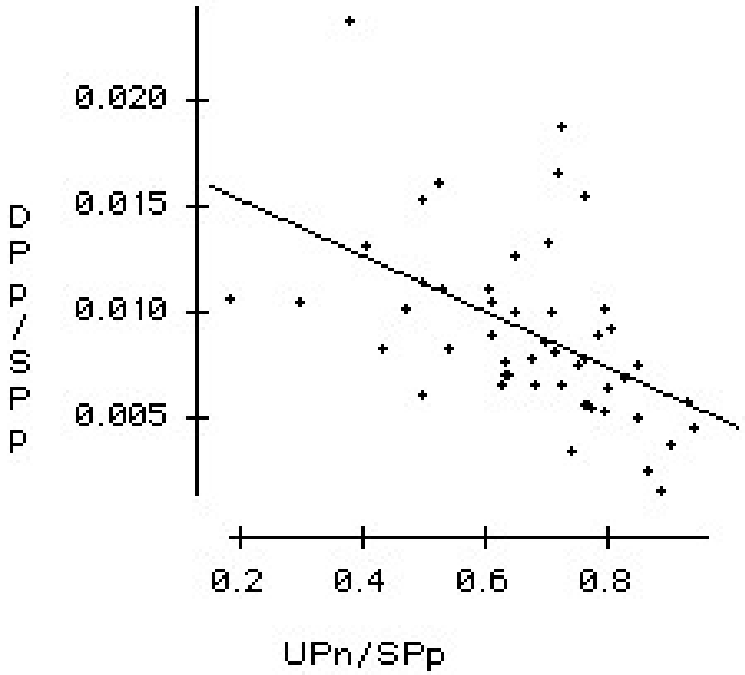
Per Capita Dorm pop vs. Urban pop by State



# Does an Urban Location Encourage Students to Live in Dorms?



## Per Capita Dorm pop vs. Urban pop by State




UPn/SPp

**Pearson Product-Moment Correlation**

No Selector

| | DPp/S_ | UPn/S_ |
|---|---|---|
| **DPp/SPp** | 1.000 | |
| **UPn/SPp** | -0.510 | 1.000 |

# Does an Urban Location Encourage Students to Live in Dorms?



Switching to a per-capita view even switches the *direction* of the association!

# Correlation Properties



$$r = \frac{1}{n-1} \Sigma \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)$$

# Correlation Properties



$$r = \frac{1}{n-1}\Sigma\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$-1 <= r <= 1$ ($= \pm 1$ only for perfect linear association)

# Correlation Properties



$$r = \frac{1}{n-1}\Sigma\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$r$ is unchanged if $x$ and $y$ are exchanged.

# Correlation Properties



$$r = \frac{1}{n-1}\Sigma\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

invariant under rescaling

# Correlation Properties



$$r = \frac{1}{n-1}\Sigma\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

Curved association and r=0 are consistent!

# Correlation Properties



$$r = \frac{1}{n-1} \Sigma \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)$$

$r$ is strongly affected by outliers.

# Correlation Properties



$$r = \frac{1}{n-1}\Sigma\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

samples from independent RV's $\Rightarrow$ $r \sim 0$

# Correlation Properties



$$r = \frac{1}{n-1}\Sigma\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

X,Y indep std normal RV's ; set $Y^* = \rho X + \sqrt{1-\rho^2}Y$ Then $(X, Y^*)$ will tend to generate data with $r \sim \rho$. (e.g. $\rho = .99$ $=>$ $\frac{\sqrt{(1-\rho^2)}}{\rho} = .14$ !)

# What Does Correlation $r$ look like?



Suppose $X$ and $Y$ are indep. std. normal RV's.

# What Does Correlation $r$ look like?



Suppose $X$ and $Y$ are indep. std. normal RV's.
Data $(x_i, y_i)$ modeled on these will tend to have $r \sim 0$.

# What Does Correlation $r$ look like?



Let $r$ be a number between $-1$ and 1.
Set $Y^* = rX + \sqrt{1 - r^2}Y$.
Then data $(x_i, y_i)$ modeled $(X, Y^*)$ will tend to have a correlation of approx. $r$.

# What Does Correlation $r$ look like?



Equally true if $Y^*$ is scaled.
So

$$(X, X + \left(\frac{\sqrt{1-r^2}}{r}\right) Y)$$

will tend to produce data with correlation $r$.

# What Does Correlation $r$ look like?



Equally true if $Y^*$ is scaled.
So

$$(X, X + \left(\frac{\sqrt{1-r^2}}{r}\right) Y)$$

will tend to produce data with correlation $r$.

| $r$ | $\frac{\sqrt{1-r^2}}{r}$ |
|---|---|
| 0 | infinite |
| .1 | 9.95 |
| .5 | 1.73 |
| .8 | .75 |
| .9 | .48 |
| .99 | .14 |

# What Does Correlation $r$ look like?



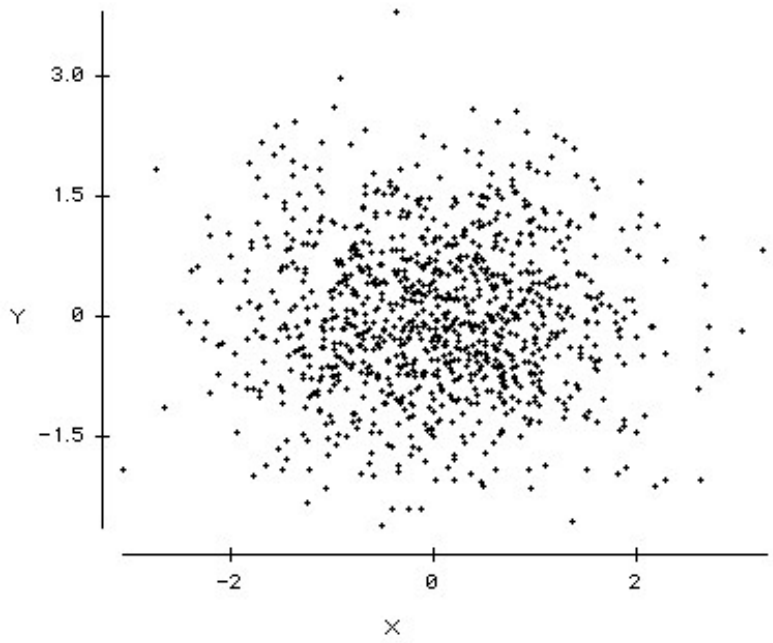
| $r$ | $\frac{\sqrt{1-r^2}}{r}$ |
|---|---|
| 0 | infinite |
| .1 | 9.95 |
| .5 | 1.73 |
| .8 | .75 |
| .9 | .48 |
| .99 | .14 |

Thus an error term $\frac{1}{7}$ the size of a perfectly associated part can still lead to a corr. coeff. of .99!

# What Does Correlation $r$ look like?



X vs. Y (1000 cases) $r \sim 0$




**Pearson Product-Moment Correlation**

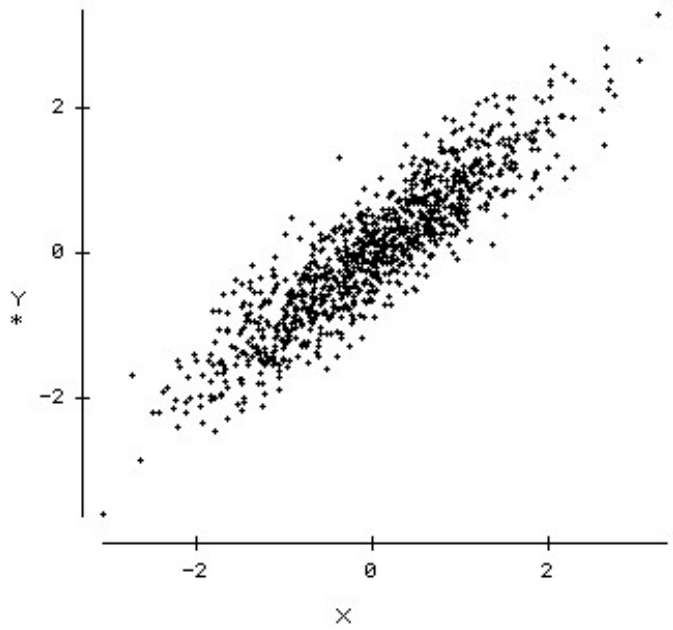
No Selector

| | Y | X |
|---|---|---|
| Y | 1.000 | |
| X | -0.015 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .9$




**Pearson Product-Moment Correlation**

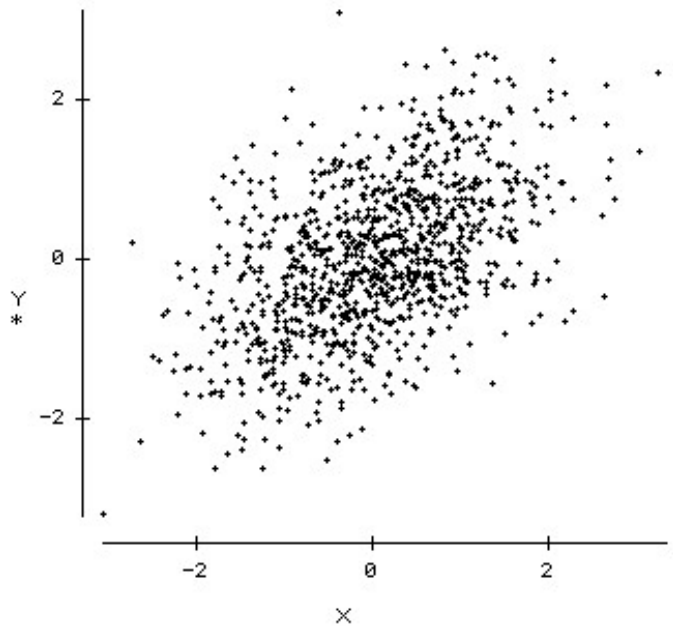
No Selector

| | Y* | X |
|---|---|---|
| Y* | 1.000 | |
| X | 0.903 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .5$




**Pearson Product-Moment Correlation**

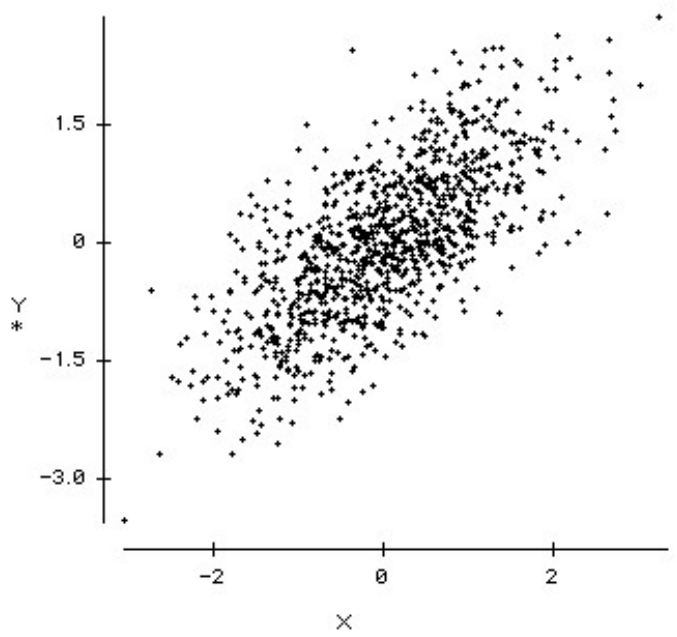
No Selector

| | Y* | X |
|---|---|---|
| **Y*** | 1.000 | |
| **X** | 0.499 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .7$




**Pearson Product-Moment Correlation**

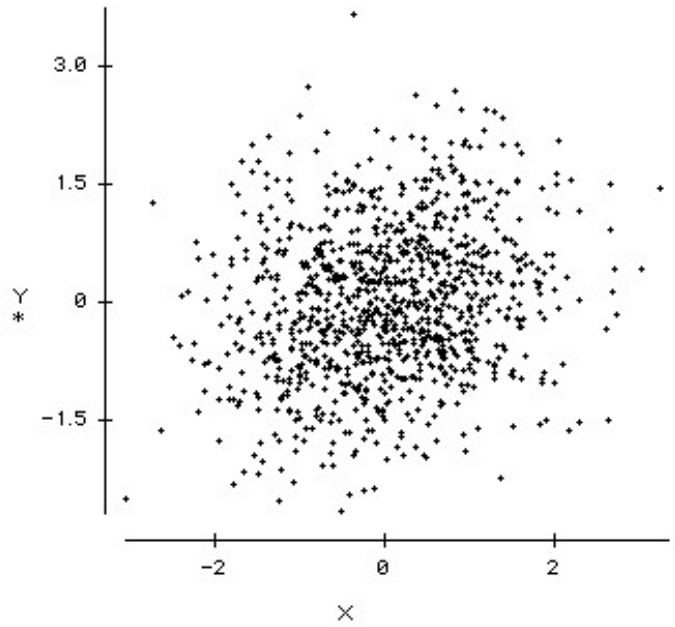
No Selector

| | Y* | X |
|---|---|---|
| **Y*** | 1.000 | |
| **X** | 0.703 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .2$




**Pearson Product-Moment Correlation**

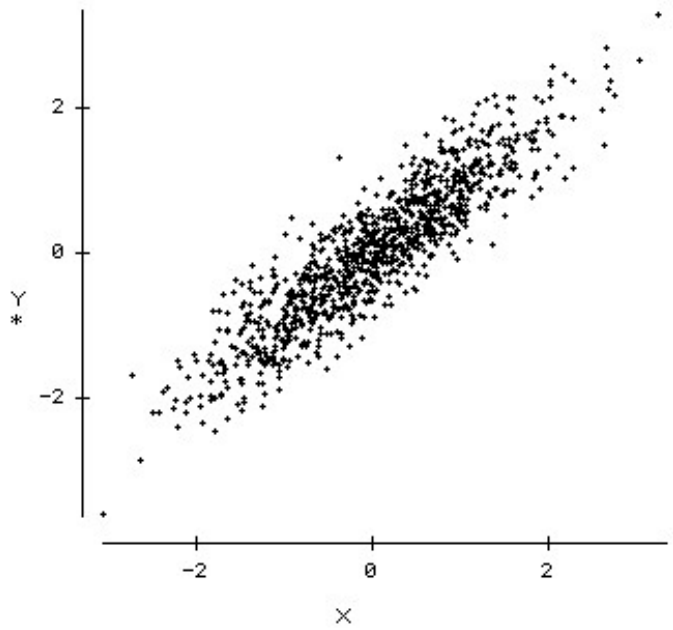
No Selector

| | Y* | X |
|---|---|---|
| **Y*** | 1.000 | |
| **X** | 0.190 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .9$




**Pearson Product-Moment Correlation**

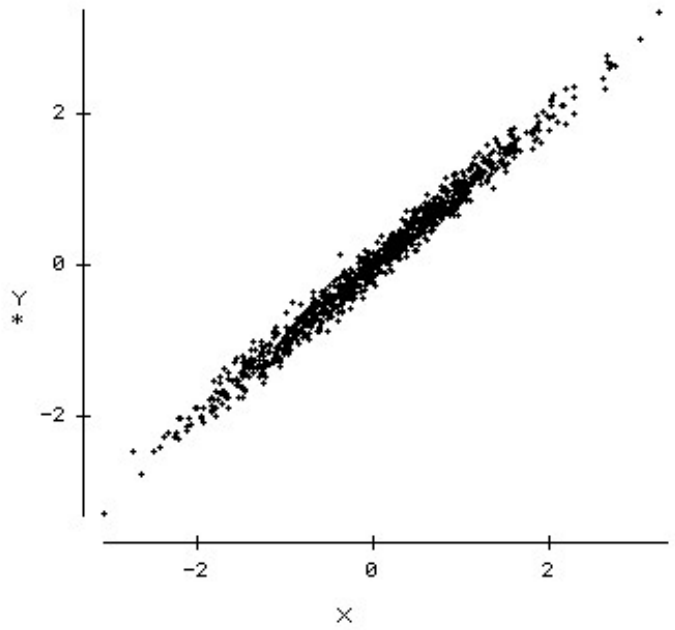
No Selector

| | Y* | X |
|---|---|---|
| Y* | 1.000 | |
| X | 0.903 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .99$




**Pearson Product-Moment Correlation**

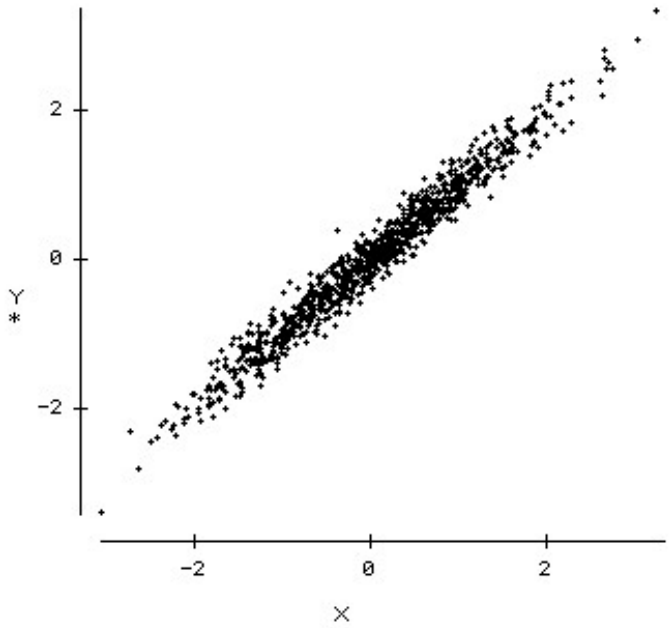
No Selector

| | Y* | X |
|---|---|---|
| **Y*** | 1.000 | |
| **X** | 0.991 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .98$




**Pearson Product-Moment Correlation**

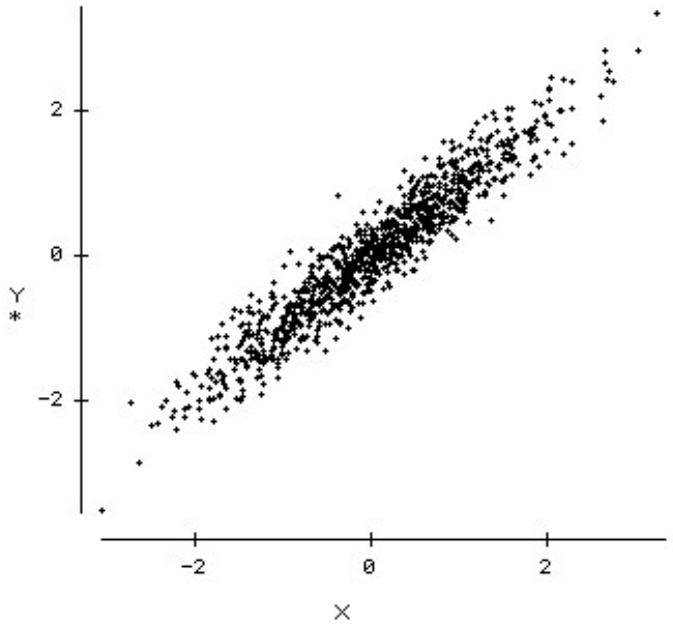
No Selector

| | Y* | X |
|---|---|---|
| Y* | 1.000 | |
| X | 0.981 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .95$




**Pearson Product-Moment Correlation**

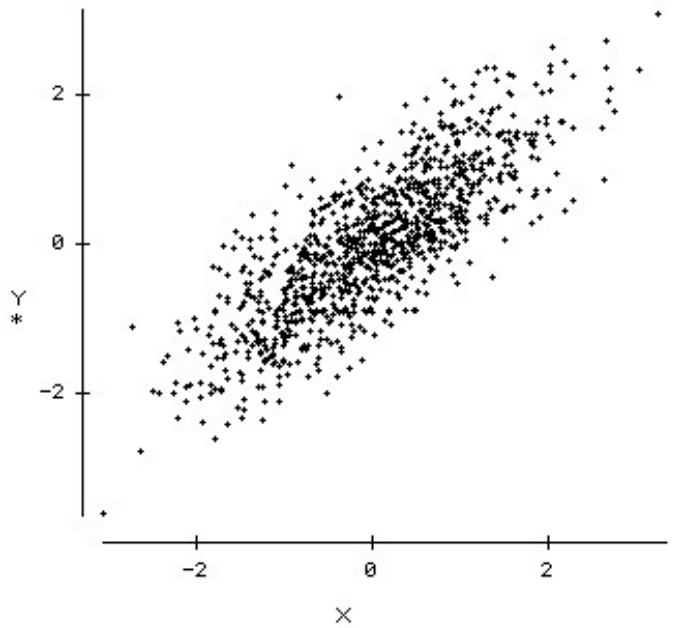
No Selector

| | Y* | X |
|---|---|---|
| **Y*** | 1.000 | |
| **X** | 0.952 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .8$




**Pearson Product-Moment Correlation**

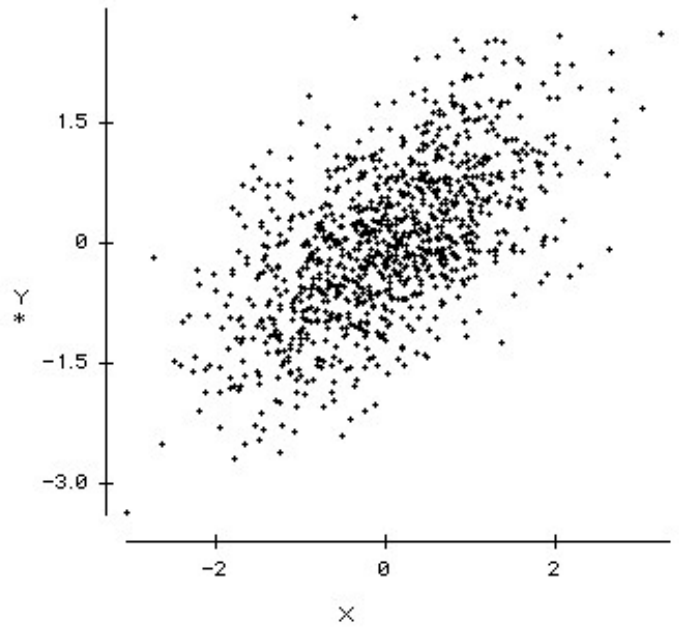
No Selector

| | Y* | X |
|---|---|---|
| Y* | 1.000 | |
| X | 0.804 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .6$




**Pearson Product-Moment Correlation**

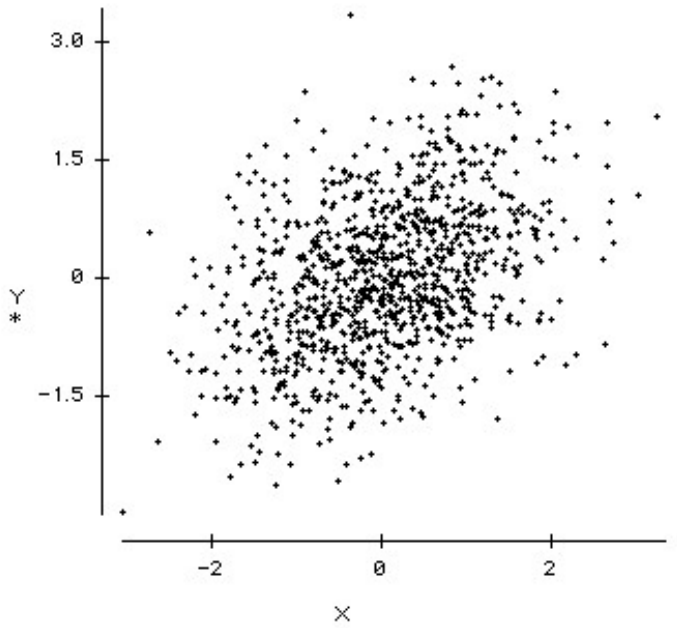
No Selector

| | Y* | X |
|---|---|---|
| **Y*** | 1.000 | |
| **X** | 0.602 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .4$




**Pearson Product-Moment Correlation**

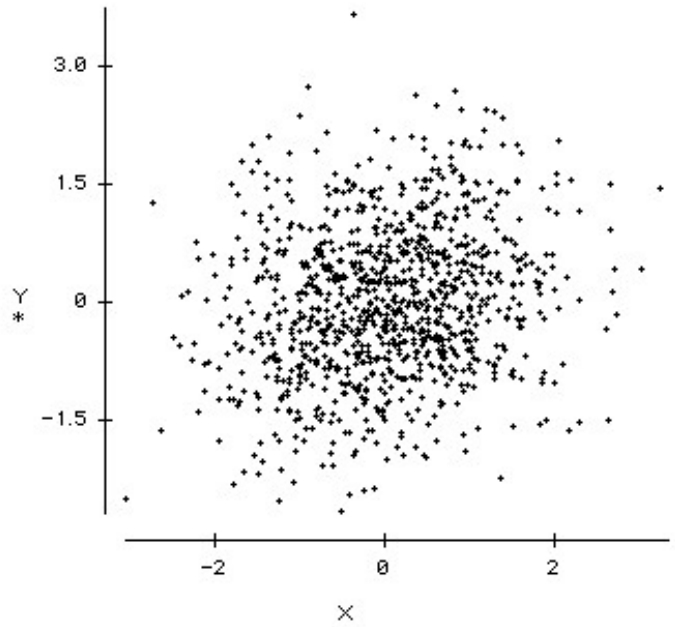
No Selector

| | Y* | X |
|---|---|---|
| Y* | 1.000 | |
| X | 0.397 | 1.000 |

# What Does Correlation $r$ look like?



X vs. $Y^*$ (1000 cases) $r \sim .2$




**Pearson Product-Moment Correlation**

No Selector

| | Y* | X |
|---|---|---|
| **Y*** | 1.000 | |
| **X** | 0.190 | 1.000 |

# Started Airfare and Distance Handout.





A good elementary example.
Worked on blackboard.
Will continue with this on Monday.