* Announcements
  - Project Step 3 due 4/2] Everyone has to schedule a meeting w/ me
  - Exams will be back Thursday!
* Pearson Correlation
* Wrap up chapter 3
* Start chapter 6] Multiple regression!

Pearson Product Moment Correlation

$\rho\{X, Y\}$ is the correlation between two random variables X and Y then

Suppose we have $X_1, \ldots, X_n$
$Y_1, \ldots, Y_n$

$$r\{X, Y\} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sum (X_i - \bar{X})^2}}$$

* will come up in IE step 3
* page 83-84 of textbook (pdf)

Concluding Chapter 3

* Skip Box-Cox transformations] Summarize different common transformations
  - Include $\log(Y_i)$, Y_i^*
  - Usually we just try $\log(Y)$

$$\sqrt{Y}$$

$$Y^2$$

* case study: we have data from a plutonium experiment
  - $n=24$ observations
  - X_i plutonium activity
  - Y_i alpha count rate

Chapter 6: Multiple Regression!

Sometimes (often) a response Y is a function of <u>multiple</u> predictors

Two Predictors

* $E\{\varepsilon_i\} = 0$
* $\sigma^2\{\varepsilon_i\} = \sigma^2$

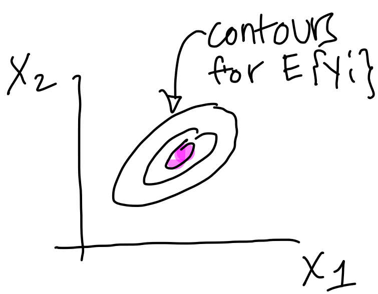
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

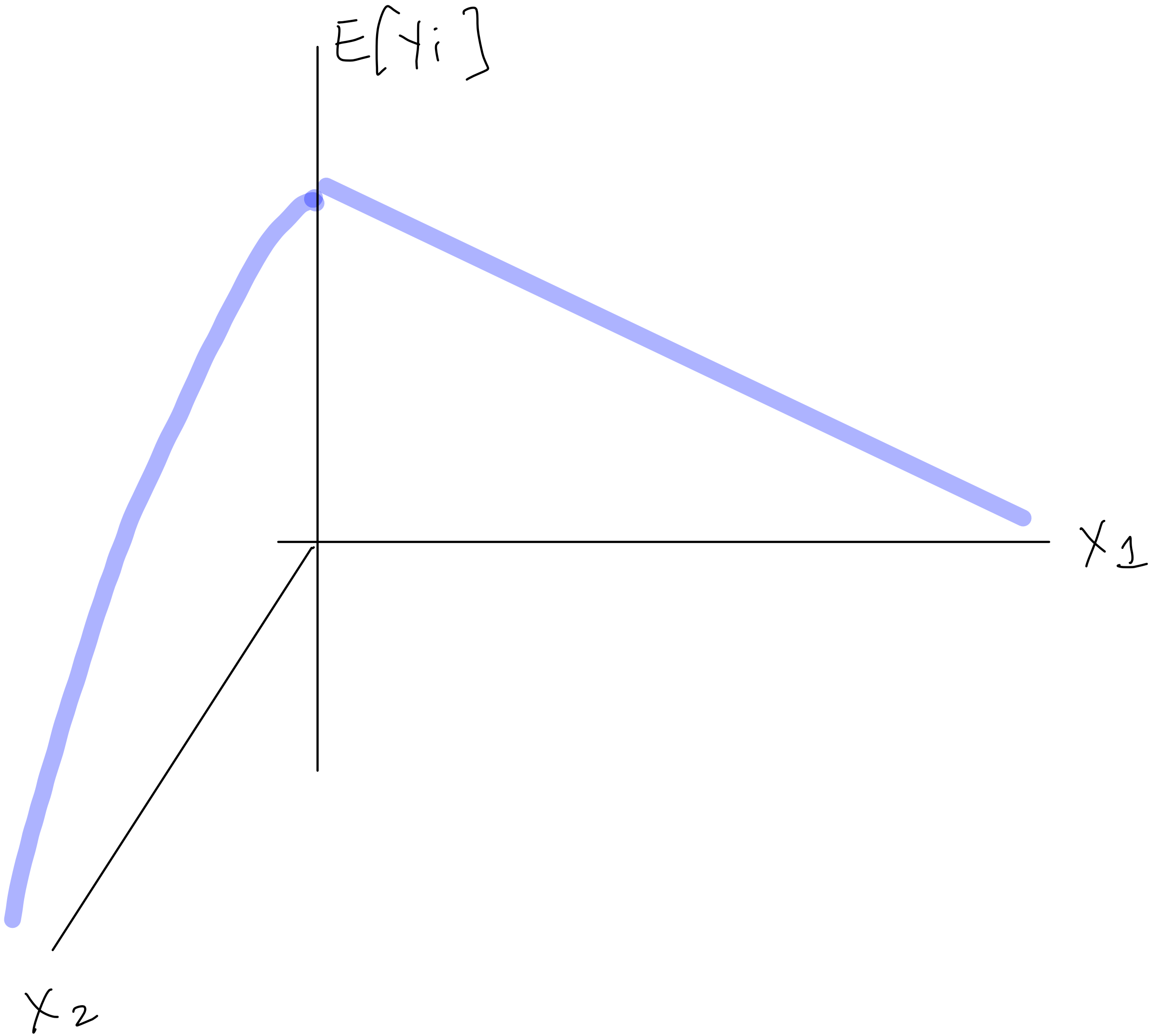
$$E\{Y_i\} = E\{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i\}$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + E\{\varepsilon_i\}$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

Regression Function or surface





E[Yi]
X1
X2

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

* $E\{\varepsilon_i\} = 0$
* $\sigma^2\{\varepsilon_i\} = \sigma^2$

Interpreting $\beta_1, \beta_2, \beta_0$

$$E\{Y_i\} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

If $X_{i1} = 0$ and $X_{i2} = 0$ then

$E\{Y_i\} = \beta_0$] If 0 is in the scope of X_1 and X_2 then β_0 can be interpreted as the expected value of Y_i at 0 covariate values

$* \, E\{\varepsilon_i\} = 0$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \qquad * \, \sigma^2\{\varepsilon_i\} = \sigma^2$$

Interpreting $\beta_1, \beta_2, \beta_0$

$$E\{Y_i\} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

Lets think about β_1,

Consider Y_j with $X_{j1} = X_{i1} + 1$,
$X_{j2} = X_{i2}$

$$E\{Y_j\} - E\{Y_i\} = \cancel{\beta_0} + \beta_1 + \beta_1 X_{i1} + \beta_2 X_{i2} - (\cancel{\beta_0} + \beta_1 X_{i1} + \beta_2 X_{i2})$$

$$= \beta_1$$

β_1 is the change in mean response per unit increase in X_1 holding X_2 constant

Multiple Linear Regression (with p-1 predictors)

$E\{\varepsilon_i\} = 0, \quad \sigma^2\{\varepsilon_i\} = \sigma^2$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

* sometimes called a "first order model" with p-1 predictors

* can be written as

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i, \quad X_{i0} = 1$$

or (later) $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

$$E\{Y_i\} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1}$$

* Interpretation of $\beta_1, \dots, \beta_{p-1}$ is analogous to before, i.e. β_1 is the mean change in response given a one unit change in X_{i1}, holding <u>all else</u> constant