Data and Regression Analysis

Lecturer: Prof. Duane S. Boning

Rev 10

Agenda

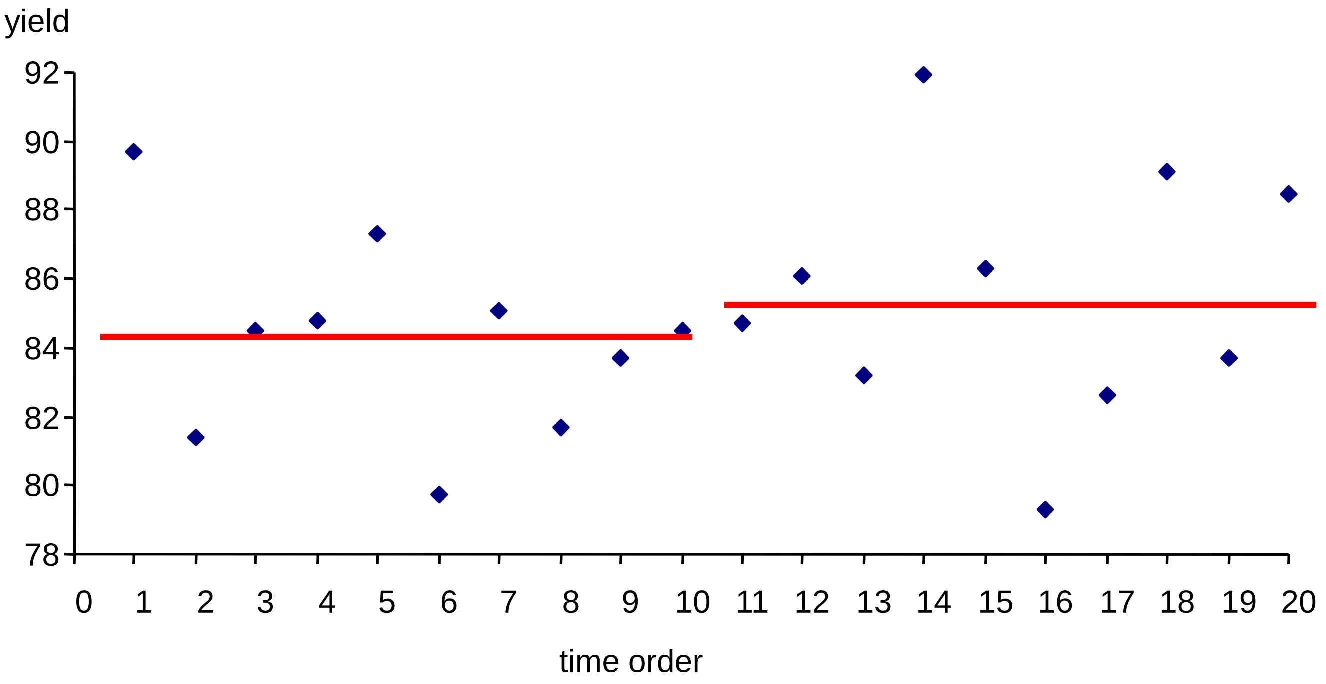
1. Comparison of Treatments (One Variable)
   - Analysis of Variance (ANOVA)

2. Multivariate Analysis of Variance
   - Model forms

3. Regression Modeling
   - Regression fundamentals
   - Significance of model terms
   - Confidence intervals

Is Process B Better Than Process A?

time order	method	yield
1	A	89.7
2	A	81.4
3	A	84.5
4	A	84.8
5	A	87.3
6	A	79.7
7	A	85.1
8	A	81.7
9	A	83.7
10	A	84.5
11	B	84.7
12	B	86.1
13	B	83.2
14	B	91.9
15	B	86.3
16	B	79.3
17	B	82.6
18	B	89.1
19	B	83.7
20	B	88.5



$$\bar{y}_A = 84.24$$
$$\bar{y}_B = 85.54$$
$$\bar{y}_B - \bar{y}_A = 1.30$$

$$H_0 : \mu_A = \mu_B$$
$$H_1 : \mu_A < \mu_B$$

Assume variances in A and B are equal.

Two Means with Internal Estimate of Variance

Method A		Method B	
count	$n_A = 10$	count	$n_B = 10$
sum	842.4	sum	855.4
average	$\bar{y}_A = 84.24$	average	$\bar{y}_B = 85.54$
sum squares	$\sum(y_A - \bar{y}_A)^2 = 75.784$	sum squares	$\sum(y_B - \bar{y}_B)^2 = 119.924$

$$\bar{y}_B - \bar{y}_A = 1.30$$

Pooled estimate of σ^2 $\quad s^2 = \frac{75.784+119.924}{10+10-2} = \frac{195.708}{18} = 10.8727$ with ν=18 d.o.f

Estimated variance of $\bar{y}_B - \bar{y}_A$

$$s^2\left(\frac{1}{n_B} + \frac{1}{n_A}\right) = \frac{2s^2}{10} = \frac{s^2}{5}$$

Estimated standard error of $\bar{y}_B - \bar{y}_A$

$$\sqrt{\frac{s^2}{5}} = \sqrt{\frac{10.8727}{5}} = 1.47$$

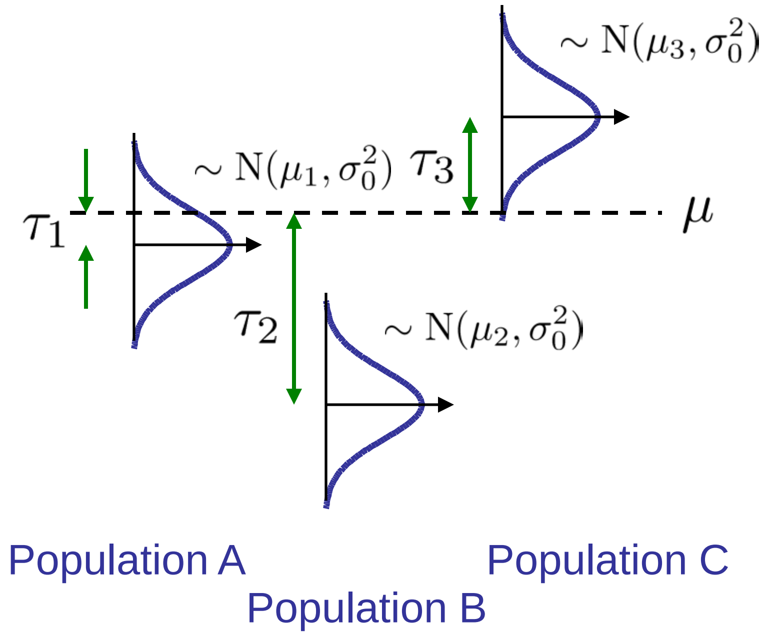
$$t_0 = \frac{(\bar{y}_B - \bar{y}_A) - (\mu_B - \mu_A)}{s\sqrt{1/n_A + 1/n_B}}$$

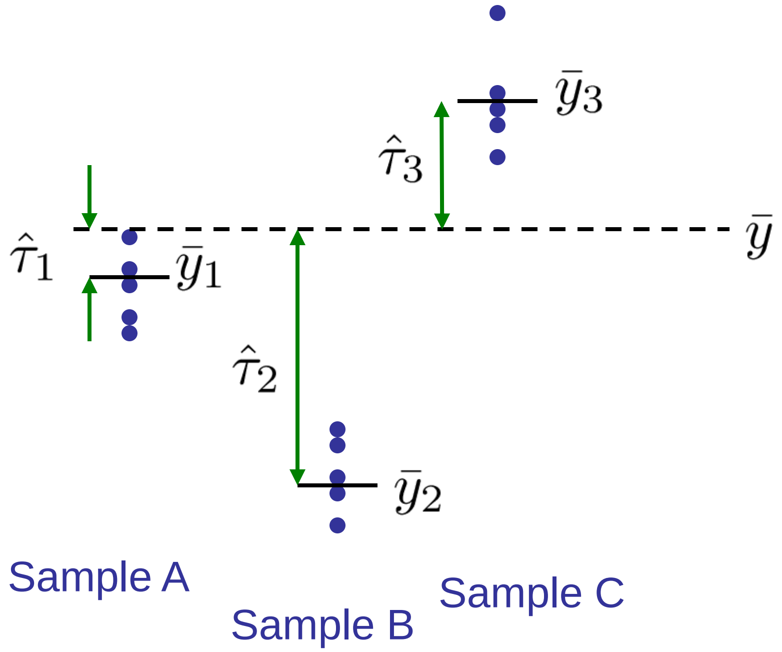
For $\mu_B - \mu_A = 0, t_0 = \frac{1.30}{1.47} = 0.88$ with $\nu = 18$ degrees of freedom.

$$\Pr(t \geq t_0) = \Pr(t \geq 0.88) = 0.195$$

So only about 80.5% confident that mean difference is “real” (significant)

Comparison of Treatments





- Consider multiple conditions (treatments, settings for some variable)
  - There is an overall mean μ and real "effects" or deltas between conditions τ_i.
  - We observe samples at each condition of interest
- Key question: are the ***observed*** differences in mean "significant"?
  - Typical assumption (should be checked): the underlying variances are all the same – usually an unknown value (σ_0^2)

Steps/Issues in Analysis of Variance

1. Within group variation
   – Estimate underlying population variance

2. Between group variation
   – Estimate group to group variance

3. Compare the two estimates of variance
   – If there is a difference between the different treatments, then the between group variation estimate will be ***inflated*** compared to the within group estimate
   – We will be able to establish confidence in whether or not observed differences between treatments are significant

   Hint: we'll be using *F* tests to look at ratios of variances

(1) Within Group Variation

- Assume that each group is normally distributed and shares a common variance σ_0^2
- SS_t = sum of square deviations within tth group (there are *k* groups)

$$SS_t = \sum_{i=1}^{n_t} (y_{ti} - \bar{y}_t)^2 \text{ where } n_t \text{ is number of samples in treatment } t$$

- Estimate of within group variance in tth group (just variance formula)

$$s_t^2 = SS_t / \nu_t = \frac{SS_t}{n_t - 1} \quad \text{where } \nu_t \text{ is d.o.f. in treatment } t$$

- Pool these (across different conditions) to get estimate of common within group variance:

$$s_R^2 = \frac{\nu_1 s_1^2 + \nu_1 s_1^2 + \cdots + \nu_k s_k^2}{\nu_1 + \nu_2 + \cdots + \nu_k} = \frac{SS_R}{\nu_R} = \frac{SS_R}{N-k} = \frac{\sum_t \sum_i (y_{ti} - \bar{y}_t)^2}{N-k} = \frac{\sum_t SS_t}{N-k}$$

- This is the within group "mean square" (variance estimate)

$$MS_R = \frac{SS_R}{\nu_R} = s_R^2$$

(2) Between Group Variation

- We will be testing hypothesis $\mu_1 = \mu_2 = \ldots = \mu_k$
- If all the means are in fact equal, then a 2nd estimate of σ^2 could be formed based on the observed differences between group means:

$$s_T^2 = \frac{\sum_{t=1}^{k} n_t(\bar{y}_t - \bar{y})^2}{k-1} = \frac{SS_T}{k-1}$$

where n_t is number of samples in treatment t, and k is the number of different treatments

- If the treatments in fact have different means, then s_T^2 estimates something larger:

$$s_T^2 \simeq \sigma_0^2 + \frac{\sum_{t=1}^{k} n_t \tau_t^2}{k-1}$$

where τ_t is the (real) difference between group t mean and the grand mean μ

Variance is "inflated" by the real treatment effects τ_t

(3) Compare Variance Estimates

- We now have two different possibilities for s_T^2, depending on whether the observed sample mean differences are “real” or are just occurring by chance (by sampling)
- Use *F* statistic to see if the ratios of these variances are likely to have occurred by chance!
- Formal test for significance:

$$\text{Reject } H_0 \ (H_0 : \text{no mean difference})$$
$$\text{if } \frac{s_T^2}{s_R^2} \text{ is significantly greater than } 1.$$

(4) Compute Significance Level

- Calculate observed *F* ratio (with appropriate degrees of freedom in numerator and denominator)
- Use *F* distribution to find how likely a ratio this large is to have occurred by chance alone
  - This is our "significance level"
  - Define observed ratio: $F_0 = s_T^2 / s_R^2$
  - If $F_0 > F_{\alpha,k-1,N-k}$
    then we say that the mean differences or treatment effects are significant to (1-α)100% confidence or better

(5) Variance Due to Treatment Effects

- We also want to estimate the sum of squared *deviations from the grand mean* among all samples:

$$SS_D = \sum_{t=1}^{k} \sum_{i=1}^{n_t} (y_{ti} - \bar{y})^2$$

$$s_D^2 = SS_D / \nu_D = \frac{SS_D}{N-1} = MS_D$$

where N is the total number of measurements

(6) Results: The ANOVA Table

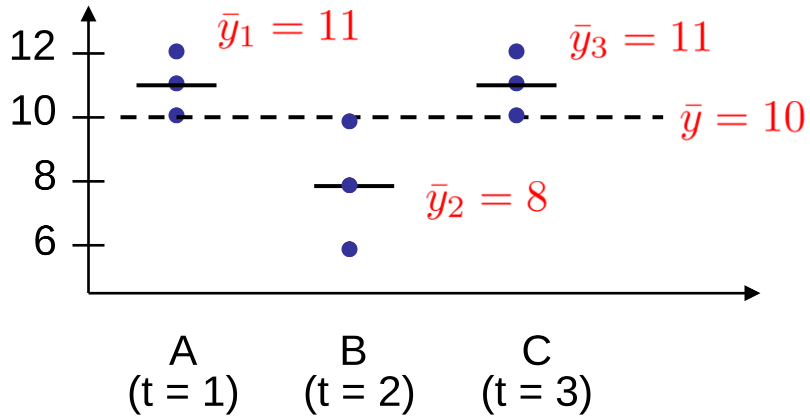
source of variation	sum of squares	degrees of freedom	mean square	F_0	$\Pr(F_0)$
Between treatments	SS_T	$k-1$	$s_T^2 = \frac{SS_T}{k-1}$	$\frac{s_T^2}{s_R^2}$	table
Within treatments	SS_R (Also referred to as "residual" SS)	$N-k$	$s_R^2 = \frac{SS_R}{N-k}$		
Total about the grand average	SS_D	$N-1$	$s_D^2 = \frac{SS_D}{N-1}$		

$SS_D = SS_T + SS_R$

$\nu_D = \nu_T + \nu_R$

Example: Anova

A	B	C
11	10	12
10	8	10
12	6	11



Excel: Data Analysis, One-Variation Anova

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
A	3	33	11	1		
B	3	24	8	4		
C	3	33	11	1		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	18	2	9	4.5	0.064	5.14
Within Groups	12	6	2			
Total	30	8				

$$F = \frac{S_T^2}{S_R^2} = \frac{9}{2} = 4.5$$

$$F_{0.05,2,6} = 5.14$$

$$F_{0.10,2,6} = 3.46$$

$$\begin{aligned} SS_1 &= (12-11)^2 + (11-11)^2 + (10-11)^2 = 2 \\ SS_2 &= 2^2 + 0^2 + 2^2 = 8 \\ SS_3 &= 1^2 + 0^2 + 1^{=}2 \end{aligned}$$

$$\begin{aligned} s_1^2 &= MS_1 = SS_1/2 = 2/2 = 1 \\ s_2^2 &= MS_2 = 8/2 = 4 \\ s_3^2 &= MS_3 = 2/2 = 1 \end{aligned}$$

$$s_R^2 = \frac{SS_1 + SS_2 + SS_3}{N-k} = \frac{12}{6} = 2$$

$$\begin{aligned} s_T^2 &= \frac{3(11-10)^2+3(8-10)^2+3(11-10)^2}{3-1} \\ &= \frac{SS_T}{\nu_T} = \frac{18}{2} = 9 \end{aligned}$$

ANOVA – Implied Model

- The ANOVA approach assumes a simple mathematical model:

$$\begin{aligned} y_{ti} &= \mu + \tau_t + \epsilon_{ti} \\ &= \mu_t + \epsilon_{ti} \end{aligned}$$

- Where μ_t is the treatment mean (for treatment type t)
- And τ_t is the treatment effect
- With ε_{ti} being zero mean normal residuals $\sim N(0,\sigma_0^2)$
- Checks
  - Plot residuals against time order
  - Examine distribution of residuals: should be IID, Normal
  - Plot residuals vs. estimates
  - Plot residuals vs. other variables of interest

MANOVA – Two Dependencies

- Can extend to two (or more) variables of interest. MANOVA assumes a mathematical model, again simply capturing the means (or treatment offsets) for each discrete variable level:

$$
\begin{array}{rcccccccc}
y_{tqi} & = & \mu & + & \tau_t & + & \beta_q & + & \epsilon_{tqi} \\
\text{^ indicates estimates:} \quad \hat{y}_{tq} & = & \hat{\mu} & + & \hat{\tau}_t & + & \hat{\beta}_q & &
\end{array}
$$

$$
\begin{array}{rcccccc}
\#\ \text{model coeffs} & = & 1 & + & k & + & n \\
 & & \uparrow & & \uparrow & & \uparrow \\
\#\ \text{independent model coeffs} & = & 1 & + & (k-1) & + & (n-1)
\end{array}
$$

Recall that our $\hat{\tau}_t$ are *not* all independent model coefficients, because $\sum \tau_t = 0$. Thus we really only have $k-1$ independent model coeffs, or $\nu_t = k-1$.

- Assumes that the effects from the two variables are ***additive***

Example: Two Factor MANOVA

- Two LPCVD deposition tube types, three gas suppliers. Does supplier matter in average particle counts on wafers?
  - Experiment: 3 lots on each tube, for each gas; report average # particles added

		Factor 1 Gas: A	B	C	
Factor 2 Tube	1	**7**	**36**	**2**	15
	2	**13**	**44**	**18**	25
		10	40	10	

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	1350.00	450.0	32.14
Error	2	28.00	14.0	Prob > F
C. Total	5	1378.00		0.0303

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Tube	1	1	150.00	10.71	0.0820
Gas	2	2	1200.00	42.85	0.0228

$$y_{tqi} = \mu + \tau_t + \beta_q + \epsilon_{tqi}$$

$$y_{tqi} = \bar{y} + (\bar{y}_t - \bar{y}) + (\bar{y}_q - \bar{y}) + (y_{tqi} - \bar{y}_t - \bar{y}_q + \bar{y})$$

$$\boxed{\begin{matrix} 7 & 36 & 2 \\ 13 & 44 & 18 \end{matrix}} \quad \boxed{\begin{matrix} 20 & 20 & 20 \\ 20 & 20 & 20 \end{matrix}} + \boxed{\begin{matrix} -10 & 20 & -10 \\ -10 & 20 & -10 \end{matrix}} + \boxed{\begin{matrix} -5 & -5 & -5 \\ 5 & 5 & 5 \end{matrix}} + \boxed{\begin{matrix} 2 & 1 & -3 \\ -2 & -1 & 3 \end{matrix}}$$

$$SS = SS_A + SS_T + SS_B + SS_R$$

MANOVA – Two Factors with Interactions

- May be interaction: not simply additive – effects may depend synergistically on both factors:

IID, ~N(0,σ^2)

$$y_{tqi} = \mu_{tq} + \epsilon_{tqi}$$

An effect that depends on both t & q factors simultaneously

t = first factor = 1,2, … k (k = # levels of first factor)
q = second factor = 1,2, … n (n = # levels of second factor)
i = replication = 1,2, … m (m = # replications at t, qth combination of factor levels

- Can split out the model more explicitly…

$$y_{tqi} = \mu + \tau_t + \beta_q + \omega_{tq} + \epsilon_{tqi}$$

Estimate by: $$\hat{y}_{tq} = \bar{y} + (\bar{y_t} - \bar{y}) + (\bar{y_q} - \bar{y}) + (\bar{y_{tq}} - \bar{y_t} - \bar{y_q} + \bar{y})$$

$$\omega_{tq} = \text{interaction effects} = (\bar{y_{tq}} - \bar{y_t} - \bar{y_q} + \bar{y})$$
$$\tau_t, \beta_q = \text{main effects}$$

MANOVA Table – Two Way with Interactions

source of variation	sum of squares	degrees of freedom	mean square	F_0	Pr(F_0)
Between levels of factor 1 (T)	SS_T	$k-1$	s_T^2	s_T^2/s_E^2	table
Between levels of factor 2 (B)	SS_B	$n-1$	s_B^2	s_B^2/s_E^2	table
Interaction	SS_I	$(k-1)(n-1)$	s_I^2	s_I^2/s_E^2	table
Within Groups (Error)	SS_E	$nk(m-1)$	s_E^2		
Total about the grand average	SS_D	$nkm-1$			

Measures of Model Goodness – R^2

- Goodness of fit – R^2
  - Question considered: how much better does the model do than just using the grand average?

$$R^2 = \frac{SS_T}{SS_D}$$

  - Think of this as the fraction of squared deviations (from the grand average) in the data which is captured by the model
- Adjusted R^2
  - For "fair" comparison between models with different numbers of coefficients, an alternative is often used

$$R^2_{\text{adj}} = 1 - \frac{SS_R/\nu_R}{SS_D/\nu_D} = 1 - \frac{s_R^2}{s_D^2}$$

  - Think of this as (1 – variance remaining in the residual). Recall $\nu_R = \nu_D - \nu_T$

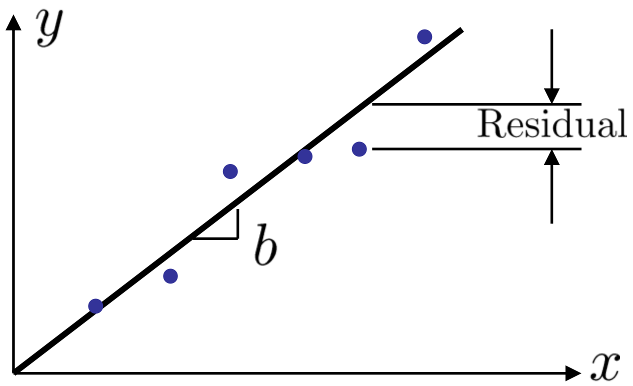
Regression Fundamentals

- Use least square error as measure of goodness to estimate coefficients in a model
- One parameter model:
  - Model form
  - Squared error
  - Estimation using normal equations
  - Estimate of experimental error
  - Precision of estimate: variance in b
  - Confidence interval for β
  - Analysis of variance: significance of b
  - Lack of fit vs. pure error
- Polynomial regression

Least Squares Regression

- We use ***least-squares*** to estimate coefficients in typical regression models
- One-Parameter Model:

$$\begin{aligned} y_i &= \beta x_i + \epsilon_i, \quad i = 1, 2, ..., n; \;\; \epsilon_i \sim N(0, \sigma^2) \\ \hat{y}_i &= b x_i \end{aligned}$$

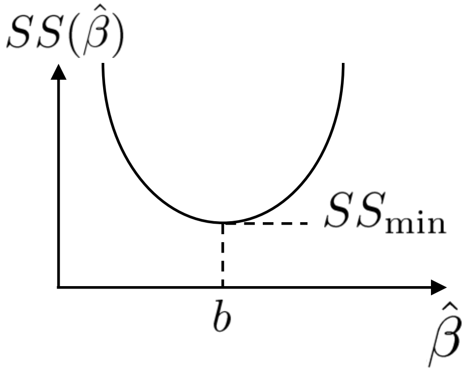


- Goal is to estimate β with "best" b
- How define "best"?
  - That b which minimizes sum of squared error between prediction and data

$$SS(\hat{\beta}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta} x_i)^2$$

  - The residual sum of squares (for the best estimate) is

$$SS_{\min} = \sum_{i=1}^{n} (y_i - b x_i)^2 = SS_R$$



Least Squares Regression, cont.

- Least squares estimation via normal equations
  - For linear problems, we need not calculate SS(β); rather, direct solution for b is possible
  - Recognize that vector of residuals will be normal to vector of x values at the least squares estimate

$$\begin{aligned} \sum (y - \hat{y})x &= 0 \\ \sum (y - bx)x &= 0 \\ \sum xy &= \sum bx^2 \end{aligned}$$

$$\Rightarrow \quad b = \frac{\sum xy}{\sum x^2}$$

- Estimate of experimental error
  - Assuming model structure is adequate, estimate s^2 of σ^2 can be obtained:

$$s^2 = \frac{SS_R}{n-1}$$

Precision of Estimate: Variance in b

- We can calculate the variance in our estimate of the slope, b:

$$\hat{V}(b) = \frac{s^2}{\sum x_i^2}$$

$$\text{s.e.}(b) = \sqrt{\hat{V}(b)}$$

$$b \pm \text{s.e.}(b)$$

- Why?

$$\begin{aligned} b &= \frac{x_1}{\sum x^2} \cdot y_1 + \frac{x_2}{\sum x^2} \cdot y_2 + \cdots \frac{x_n}{\sum x^2} \cdot y_n \\ &= a_1 y_1 + a_2 y_2 + \cdots + a_n y_n \end{aligned}$$

$$\begin{aligned} V(b) &= (a_1^2 + a_2^2 + \cdots + a_n^2)\sigma^2 \\ &= \left[\left(\frac{x_1}{\sum x^2}\right)^2 + \cdots + \left(\frac{x_n}{\sum x^2}\right)^2 \right] \sigma^2 \\ &= \frac{\sum x^2}{(\sum x^2)^2} \sigma^2 \\ &= \frac{\sigma^2}{\sum x^2} \end{aligned}$$

Confidence Interval for β

- Once we have the standard error in b, we can calculate confidence intervals to some desired (1-α)100% level of confidence

$$\frac{b-\beta}{\text{s.e.}(b)} \sim t \qquad \Rightarrow \qquad \beta = b \pm t_{\alpha/2} \cdot \text{s.e.}(b)$$

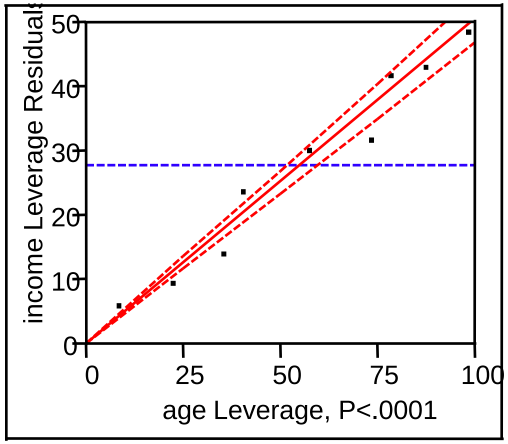
- Analysis of variance
  - Test hypothesis: $H_0 : \beta = b = 0$
  - If confidence interval for β includes 0, then β not significant
  - Degrees of freedom (need in order to use t distribution)

$$\begin{array}{ccccc} \sum y_i^2 & = & \sum \hat{y}_i^2 & + & \sum (y_i - \hat{y}_i)^2 \\ n & = & p & + & n-p \end{array}$$

p = # parameters estimated
by least squares

Example Regression

Age	Income
8	6.16
22	9.88
35	14.35
40	24.06
57	30.34
73	32.17
78	42.18
87	43.23
98	48.76



Whole Model

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	8836.6440	8836.64	1093.146
Error	8	64.6695	8.08	Prob > F
C. Total	9	8901.3135		<.0001

Tested against reduced model: Y=0

Parameter Estimates

Term		Estimate	Std Error	t Ratio	Prob>|t|
Intercept	Zeroed	0	0	.	.
age		0.500983	0.015152	33.06	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
age	1	1	8836.6440	1093.146	<.0001

- Note that this simple model assumes an intercept of zero – model must go through origin
- We will relax this requirement soon

Lack of Fit Error vs. Pure Error

- Sometimes we have replicated data
  - E.g. multiple runs at same x values in a designed experiment

- We can decompose the residual error contributions

$$SS_R = SS_L + SS_E$$

Where
SS_R = residual sum of squares error
SS_L = lack of fit squared error
SS_E = pure replicate error

- This allows us to TEST for lack of fit
  - By "lack of fit" we mean evidence that the linear model form is inadequate

$$\frac{s_L^2}{s_E^2} \sim F_{\nu_L,\nu_E}$$

Regression: Mean Centered Models

- Model form $y = \alpha + \beta(x - \bar{x})$
- Estimate by $\hat{y} = a + b(x - \bar{x}), \quad (y_i - \hat{y}_i) \sim \mathrm{N}(0, \sigma^2)$

Minimize $SS_R = \sum_{i=1}^{k}(y_i - \hat{y}_i)^2$ to estimate α and β

$$a = \bar{y}$$

$$\mathrm{E}(a) = \alpha$$

$$\mathrm{Var}(a) = \mathrm{Var}\left[\frac{\sum y_i}{k}\right] = \frac{\sigma^2}{k}$$

$$b = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$\mathrm{E}(b) = \beta$$

$$\mathrm{Var}(b) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}$$

Regression: Mean Centered Models

- Confidence Intervals

$$
\begin{aligned}
\hat{y}_i &= \bar{y} + b(x_i - \bar{x}) \\
\\
\mathrm{Var}(\hat{y}_i) &= \mathrm{Var}(\bar{y}) + (x_i - \bar{x})^2 \mathrm{Var}(b) \\
&= \frac{s^2}{n} + \frac{s^2 (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = s^2_{\hat{y}_i}
\end{aligned}
$$

- Our confidence interval on y widens as we get further from the center of our data!

$$\hat{y}_i \pm t_{\alpha/2} \cdot s_{\hat{y}_i}$$

Polynomial Regression

- We may believe that a higher order model structure applies. Polynomial forms are also linear in the coefficients and can be fit with least squares

$$\eta = \beta_0 + \beta_1 x + \beta_2 x^2$$

Curvature included through x^2 term

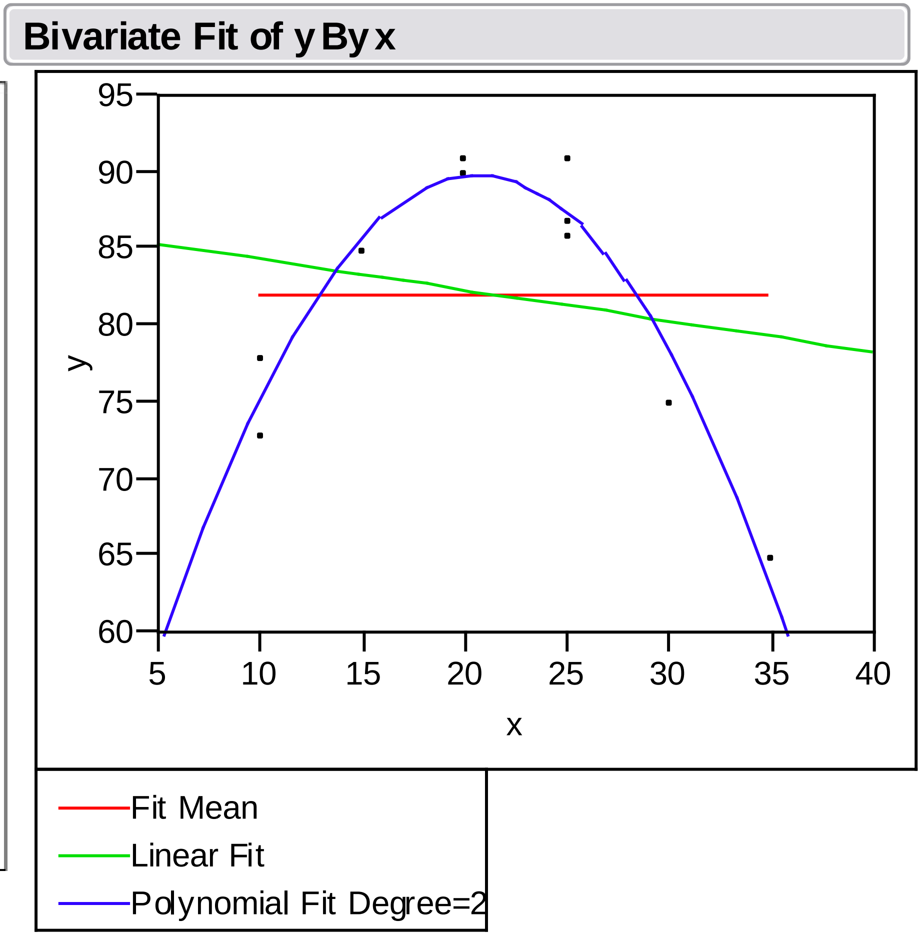
- Example: Growth rate data

Regression Example: Growth Rate Data

Growth Rate Data

Observation Number	Amount of Supplement (Grams) x	Growth Rate (Coded units) y
1	10	73
2	10	78
3	15	85
4	20	90
5	20	91
6	25	87
7	25	86
8	25	91
9	30	75
10	35	65

Image by MIT OpenCourseWare.



- Replicate data provides opportunity to check for lack of fit

Growth Rate – First Order Model

- Mean significant, but linear term not
- Clear evidence of lack of fit

Source	Sum of squares	Degrees of freedom	Mean square
Model	$S_M = 67{,}428.6 \begin{cases} \text{mean } 67{,}404.1 \\ \text{extra for linear } 24.5 \end{cases}$	$2 \begin{cases} 1 \\ 1 \end{cases}$	67,404.1 24.5
Residual $\begin{cases} \text{lack of fit} \\ \text{pure error} \end{cases}$	$S_R = 686.4 \begin{cases} S_L = 659.40 \\ S_E = 27.0 \end{cases}$	$8 \begin{cases} 4 \\ 4 \end{cases}$	$85.8 \begin{cases} 164.85 \\ 6.75 \end{cases}$ ratio = 24.42
Total	$S_T = 68{,}115.0$	10	

Image by MIT OpenCourseWare.

Growth Rate – Second Order Model

- No evidence of lack of fit
- Quadratic term significant

Source	Sum of squares	Degrees of freedom	Mean square
Model	$S_M = 68{,}071.8 \begin{cases} \text{mean } 67{,}404.1 \\ \text{extra for linear } 24.5 \\ \text{extra for quadratic } 643.2 \end{cases}$	$3 \begin{cases} 1 \\ 1 \\ 1 \end{cases}$	67,404.1 24.5 643.2
Residual	$S_R = 43.2 \begin{cases} S_L = 16.2 \\ S_E = 27.0 \end{cases}$	$7 \begin{cases} 3 \\ 4 \end{cases}$	$\begin{cases} 5.40 \\ 6.75 \end{cases}$ ratio = 0.80
Total	$S_T = 68{,}115.0$	10	

Image by MIT OpenCourseWare.

Polynomial Regression In Excel

- Create additional input columns for each input
- Use “Data Analysis” and “Regression” tool

x	x^2	y
10	100	73
10	100	78
15	225	85
20	400	90
20	400	91
25	625	87
25	625	86
25	625	91
30	900	75
35	1225	65

Regression Statistics	
Multiple R	0.968
R Square	0.936
Adjusted R Square	0.918
Standard Error	2.541
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	665.706	332.853	51.555	6.48E-05
Residual	7	45.194	6.456		
Total	9	710.9			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	35.657	5.618	6.347	0.0004	22.373	48.942
x	5.263	0.558	9.431	3.1E-05	3.943	6.582
x^2	-0.128	0.013	-9.966	2.2E-05	-0.158	-0.097

Polynomial Regression

- Generated using JMP package

Analysis of Variance

Source	DF	Sum of Square	Mean Squar	F Ratio
Model	2	665.70617	332.853	51.5551
Error	7	45.19383	6.456	Prob > F
C. Total	9	710.90000		<.0001

Lack Of Fit

Source	DF	Sum of Square	Mean Squar	F Ratio
Lack Of Fit	3	18.193829	6.0646	0.8985
Pure Error	4	27.000000	6.7500	Prob > F
Total Error	7	45.193829		0.5157
				Max RSq
				0.9620

Summary of Fit

RSquare	0.936427
RSquare Adj	0.918264
Root Mean Sq Error	2.540917
Mean of Response	82.1
Observations (or Sum Wgts)	10

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	35.657437	5.617927	6.35	0.0004
x	5.2628956	0.558022	9.43	<.0001
x*x	-0.127674	0.012811	-9.97	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
x	1	1	574.28553	88.9502	<.0001
x*x	1	1	641.20451	99.3151	<.0001

Summary

- Comparison of Treatments – ANOVA
- Multivariate Analysis of Variance
- Regression Modeling

Next Time

- Time Series Models
- Forecasting