Ch 2

- $b_1 \sim N\left(\beta_1, \frac{\sigma^2}{\Sigma(x_i-\bar{x})^2}\right)$
  - thus $\frac{b_1-\beta_1}{\sigma(b_1)} \sim N(0,1)$ * Although it isn't often that we know $\sigma(b_1)$ so we estimate
  - and $\frac{b_1-\beta_1}{se(b_1)} \sim t_{n-2}$
- $(1-\alpha)$ Confidence Interval for β_1 : $b_1 \pm t(1-\alpha/2, n-2)$
  ↑ Add clb to model statement in Reg or GLM : 'model Y = x_i / clb;'

Two-Sided Test ① $H_0: \beta_1 = 0$
$H_a: \beta_1 \neq 0$
* this test is given in proc reg | proc glm
* Reject when $\left|\frac{b_1}{se(b_1)}\right| > t(1-\alpha/2; n-2)$

One-Sided Test ② $H_0: \beta_1 \leq 0$
$H_a: \beta_1 > 0$
* Reject when $\frac{b_1}{se(b_1)} > t(1-\alpha; n-2)$

or

$H_0: \beta_1 \geq 0$
$H_a: \beta_1 < 0$
* Reject when $\frac{b_1}{se(b_1)} < t(1-\alpha; n-2)$

- $b_0 \sim N\left(\beta_0, \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\Sigma(x_i-\bar{x})^2}\right]\right)$
  - thus $\frac{b_0-\beta_0}{\sigma(b_1)} \sim N(0,1)$
  - and $\frac{b_0-\beta_0}{se(b_1)} \sim t_{n-2}$
- $(1-\alpha)$ Confidence Interval for β_0: $b_0 \pm t(1-\alpha/2, n-2)$
  ↑ Add clb to model statement in Reg or GLM : 'model Y = x_i / clb;'
- Follows same one-sided and twosided t-tests as b_1 above

- Sampling Distribution $\hat{Y}_h = b_0 + b_1 X_h$

  $\hat{Y}_h \sim N\left(\beta_0 + \beta_1 X_h,\ \sigma\left[\frac{1}{n}\ \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right)\right.$

  thus, $\dfrac{\hat{Y}_h - \cancel{E}(\hat{Y}_h)}{\sigma(\hat{Y}_h)} \sim N(0,1)$

  and, $\dfrac{\hat{Y}_h - E(\hat{Y}_h)}{se(\hat{Y}_h)} \sim t_{(n-2)}$ for $y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

CI of mean
- $(1-\alpha)$ Confidence Interval $E(Y_h)$: $\hat{Y}_h \pm t(1-\alpha/2, n-2)\, se(\hat{Y}_h)$

  ↑ add clm to model statement in reg/GLM: 'model Y=x/clm;'

CI of prediction
- $(1-\alpha)$ Confidence Interval for $\hat{Y}_{h\,new}$

  w/ known parameters: $E\{Y_h\} \pm z\left(1-\frac{\alpha}{2}\right)\sigma$

  w/ unknown parameters: $\hat{Y}_h \pm t\left(1-\frac{\alpha}{2}\right) s\{\text{prediction}\}$

  ↑ add cli to mode statement in reg/GLM: 'model Y=x /cli;'

CI of mean of pred
- $(1-\alpha)$ Confidence Interval for $\overline{Y}_{h\,new}$: $\hat{Y}_h \pm t\left(1-\frac{\alpha}{2}, n-2\right) s\{\text{pred mean}\}$

Sum of Squares

$SSTO = \sum(Y_i - \bar{Y})^2$ = total variation — $n-1$ degrees of freedom

$SSE = \sum(Y_i - \hat{Y}_i)^2$ = variation unexplained — $n - \#$ of coeff. degrees of freedom

$SSR = \sum(\hat{Y}_i - \bar{Y}_i)^2$ = variation explained — $n-1-(n-\#\text{ coeff})$ degrees of freedom

$SSTO = SSE + SSR$

Mean Squares

$MSR = \dfrac{SSR}{n-1-(n-\#\text{ of coeff})}$ $\quad E\{MSR\} = \sigma^2 + \beta_1^2 \sum(X_i - \bar{X})^2$

$MSE = \dfrac{SSE}{n - \#\text{ of coeff}}$ $\quad E\{MSE\} = \sigma^2$

*Sum of Squares and mean squares are reported in proc reg/proc glm

<u>F-test</u> $H_0: B_1 = 0$

$H_a: B_1 \neq 0$

$$F^* = \frac{MSR}{MSE} = \frac{\frac{SSR}{\sigma^2}}{SSRdoF} \div \frac{\frac{SSE}{\sigma^2}}{SSEdoF}$$

* this can be done by adding a test statement to proc reg; 'test x=0';

Reject $F^* > F(1-\alpha, SSRdoF, SSEdoF)$

<u>Full Model</u> $Y_i = \beta_0 + \beta_1 X_1 + \epsilon_i$

- We can find a reduced model by testing coefficients = 0

<u>Coefficent of Determination</u> $= R^2 = 1 - \frac{SSE}{SSTO} = \frac{SSR}{SSTO}$ * Shows the proportion of variation explained by the <u>LINEAR</u> model

<u>Coefficient of correlation</u> $= r = \pm\sqrt{R^2}$

<u>Bivariate Distribution</u> $Y_1 \sim N(\mu_1, \sigma_1^2)$ $Y_2 \sim N(\mu_2, \sigma_2^2)$

$$f(Y_1, Y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \cdot e^{\left\{-\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{Y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho_{12}\left(\frac{Y_1-\mu_1}{\sigma_1}\right)\left(\frac{Y_2-\mu_2}{\sigma_2}\right) + \left(\frac{Y_2-\mu_2}{\sigma_2}\right)^2\right]\right\}}$$

- where ρ_{12} is the <u>coefficient of correlation</u> between Y_1 and Y_2

$$\rho_{12} = \frac{\sigma_{12}}{\sigma_1\sigma_2} = \frac{cov(Y_1, Y_2)}{var(Y_1)\, var(Y_2)}$$

<u>Conditional Probability distribution</u>

$$f(Y_1|Y_2) = \frac{f(Y_1, Y_2)}{f(Y_2)}$$

Point estimator of ρ_{12}: is denoted r_{12}

$$r_{12} = \frac{\Sigma(Y_{i1}-\bar{Y}_1)(Y_{i2}-\bar{Y}_2)}{\left[\Sigma(Y_{i1}-\bar{Y}_1)^2\,\Sigma(Y_{i2}-\bar{Y}_2)^2\right]^{1/2}}$$

(pearson product moment correlation coefficient for bivariate normal)

Test $H_0: \rho_{12} = 0$

$H_0: \rho_{12} \neq 0$

$$t^* = \frac{r_{12}\sqrt{n-2}}{\sqrt{1-r_{12}^2}}$$

Reject if $|t^*| > t(1-\alpha/2, n-2)$

CODE:
```
proc corr pearson fisher(alpha=α);
    var x1 x2;
run;
```
Note fisher gives confidence intervals

OUTPUT

Variable	N	Mean	St. Dev	Sum	Min	Max
X_1						
X_2						

	X_1	X_2
X_1	ρ_{11}	ρ_{12} P value for $H_0: \rho_{12}=0$
X_2	ρ_{12} P value for $H_0: \rho_{12}=0$	ρ_{22}

Variable	w/ Variable	N	Sample Corr	Fisher's Z	Bias adj	Corr est	CI LB	CI UB	pvalue pro
X_1	X_2	n							

Spearman Rank Coefficient for Y_1, Y_2 considerably different from the bivariate normal

1) Rank $Y_{11}, \ldots, Y_{n1}$ from 1 to n. (denote Rank of Y_{i1} as R_{i1})

2) $$r_s = \frac{\Sigma(R_{i1} - \bar{R}_1)(R_{i2} - \bar{R}_2)}{[\Sigma(R_{i1} - \bar{R}_1)^2 \Sigma(R_{i2} - \bar{R}_2)^2]^{1/2}}$$

SAS Code:

```
proc corr spearman
    var x1 x2;
run;
```

OUTPUT:

Variable	N	Mean	Std Dev	Median	Min	Max
X_1	n					
X_2	n					

	X_1	X_2
X_1	$r_{s_{11}}$	r_{s12} p value $r_{s_{12}}=0$
X_2	r_{s21} p value $r_{s_{12}}=0$	r_{s22}