

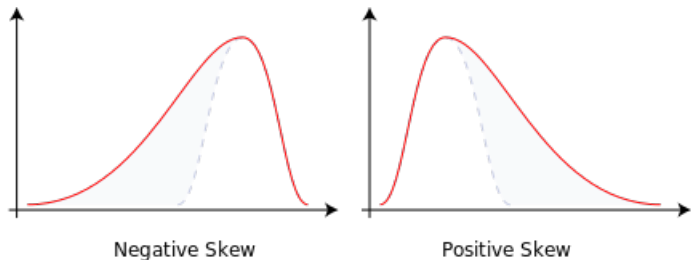
# STA2601 Applied Statistics II

## Chapter 1

*Integral* is the distribution function, or cumulative probability function,  $F_X(x)$

*Derivative* is the probability function, or probability density function  $f_x(x)$

$$Z = \frac{X - \mu}{\sigma}$$



Exponential distribution:  $\frac{1}{k}e^{-\frac{x}{k}}$  parameter is  $k$ .

$$\mu = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

$$\sigma^2 = E(X - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y]. \end{aligned}$$

## Study Unit 2

→ Unbiased estimator  $E(T) = \theta$

→  $\text{Var}(cX) = c^2 \text{Var}(X)$

→ Most efficient estimator has the smallest variance.

Least squares estimation:

$$1) Q(\theta_1, \dots, \theta_k) = \sum_{i=1}^n (X_i - E(X_i))^2$$

2) Replace  $E(X_i)$  with given  $\theta$

3) Take partials  $\frac{\partial Q}{\partial \theta_j}$

4) Set partials = 0 and solve for each  $\hat{\theta}_i$  then solve simultaneousness equations.

Maximum likelihood estimation: (MLE)

$$1) \text{ Get likelihood function } L(\theta) = f_X(X_1, \theta) f_X(X_2, \theta) \dots f_X(X_n, \theta) = \prod_{i=1}^n f_X(X_i, \theta)$$

2) Take log of likelihood function if needed

3) Take partials  $\frac{\partial Q}{\partial \theta_j}$

4) Set partials = 0 and solve for  $\hat{\theta}$  ! Remember hat ^

		Decision based on the data	
		Do not reject $H_0$	Reject $H_0$
The true state of nature	$H_0$ is true	Good decision	Type I error $\alpha$
	$H_1$ is true	Type II error $\beta$	Good decision

Significance level  $\alpha = P(H_0 \text{ is rejected} | H_0 \text{ is true})$

Power of the test  $1 - \beta = P(\text{not rejecting } H_0 | H_1 \text{ is true})$

Confidence level  $1 - \alpha$

- If this  $p$ -value is **very small**,  $\bar{x}$  is said to be **highly significant** (usually if  $p \ll \alpha$ ).
- If the  $p$ -value is **fairly small**,  $\bar{x}$  is said to be **significant** (usually if  $p < \alpha$ ).
- If the  $p$ -value is **large**,  $\bar{x}$  is said to be **not significant** (usually if  $p > \alpha$ ).

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 = \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{\sigma} \right)^2$$

$$\text{Var}(X) = E(X_i^2) - [E(X_i)]^2$$

$$\bar{X} \sim n\left(\theta, \frac{\sigma^2}{n}\right) \quad Z = \frac{\sqrt{n}(\bar{X} - \theta)}{\sigma}$$

#### Theorem 1.4

Let  $U$  and  $V$  be independent variates such that  $U \sim n(0; 1)$  and  $V \sim \chi_d^2$  and let

$$T = \frac{U}{\sqrt{V/d}}. \text{ Then } T \sim t_d.$$

$$E(X_i^2) = \theta + \mu^2 \quad E(\bar{X}^2) = \frac{\theta}{n} + \mu^2$$

$$\text{Var}(X_i) = E(X_i^2) - \mu^2$$

$$\begin{aligned} \Rightarrow E(X_i^2) &= \text{Var}(X_i) + \mu^2 \\ &= \theta + \mu^2 \end{aligned}$$

$$\text{Var}(\bar{X}) = E(\bar{X}^2) - \mu^2$$

$$\begin{aligned} \Rightarrow E(\bar{X}^2) &= \text{Var}(\bar{X}) + \mu^2 \\ &= \frac{\theta}{n} + \mu^2 \end{aligned}$$

## Study Unit 4

Van der Waerden's formula:  $\frac{r_i}{n+1}$

$\chi^2$  Goodness-of-fit test:

$$Y^2 = \sum_{i=1}^k \frac{(N_i - n\pi_i)^2}{n\pi_i}$$

$k$  number of intervals

$N_i$  is then number of items falling into the interval calculated from  $X_i = \sigma Z + \mu$

$n\pi_i$  is interval weighting  $\times$  number of observations.

## If $n\pi_i < 5$ then pool two or more cells

$Y^2 \sim \chi_{k-1}^2$  if distribution is fully specified, otherwise

$Y^2 \sim \chi_{k-1-r}^2$  where  $r$  is number of unknown parameters.

Hypotheses:

$H_0$ : The sample comes from a  $n(\mu, \sigma^2)$  distribution

$H_1$ : The sample does not come from a  $n(\mu, \sigma^2)$  distribution

Reject null hypothesis if  $Y^2 > \chi_{k-1-r}^2$

Compute the MLEs based on the ungrouped data. Compute  $Y^2$  as before.

If  $Y^2 < \chi_{\alpha; k-r-1}^2$  : do not reject  $H_0$

If  $Y^2 > \chi_{\alpha; k-1}^2$  : reject  $H_0$

If  $\chi_{\alpha; k-r-1}^2 < Y^2 < \chi_{\alpha; k-1}^2$  : decision uncertain

$$\mu \text{ known: } \hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \mu)^2$$

$$\sigma^2 \text{ known: } \hat{\mu} = \frac{1}{n} \sum X_i = \bar{X}$$

$$\mu \text{ and } \sigma^2 \text{ unknown: } \hat{\mu} = \frac{1}{n} \sum X_i = \bar{X}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$$

Use above when testing grouped data from a normal distribution because MLE's are difficult to compute.

## Method of moments test for normality

**Must perform two hypotheses tests**

Skewness

$$\beta_1 = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}}$$

Hypotheses:

$H_0$ :  $B_1 = 0$  i.e. symmetric

→ One-sided  $H_1$ :  $B_1 < 0$  Reject if  $B_1 < -TableValue$  or  $H_1$ :  $B_1 > 0$  Reject if  $B_1 > TableValue$ . 5% level

→ Two-sided  $H_1$ :  $B_1 \neq 0$  Reject if  $|B_1| > TableValue$ . 10% level

Kurtosis

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$$

Hypotheses:

$H_0$ :  $B_2 = 3$  i.e. from normal distribution

$B_2 > 3$  leptokurtic

$B_2 < 3$  platykurtic

If the alternative is  $\beta_2 < 3$ , reject  $H_0$  at the 5% level if  $B_2 < \text{lower 5\% point in table B}$ .

If the alternative is  $\beta_2 > 3$ , reject  $H_0$  at the 5% level if  $B_2 > \text{upper 5\% point in table B}$ .

If the alternative is  $\beta_2 \neq 3$ , reject  $H_0$  at the 10% level if  $B_2 < \text{lower 5\% point}$  or if  $B_2 > \text{upper 5\% point in table B}$ .

Use **standardised mean deviation** to test kurtosis if sample size is smaller than 50, statistic  $A$

Same hypotheses as  $B_2$ , 4th moment test.

$$\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$$

## Study Unit 5

### Contingency Table Analysis

Fixed Grand total:

The null hypothesis of independence is

$$H_0 : \pi_{ij} = \pi_{i.}\pi_{.j}; \quad i = 1, \dots, h; \quad j = 1, \dots, k.$$

Fixed row or column totals:

The null hypothesis of independence is that the probability of falling into category  $i$  is the same for all  $k$  populations:

$$H_0 : \pi_{i1} = \pi_{i2} = \dots = \pi_{ik} \quad \text{for } i = 1, \dots, h.$$

Test Statistic:

$$e_{ij} = \frac{N_{i.}N_{.j}}{N_{..}}$$

$$Y^2 = \sum_{i=1}^h \sum_{j=1}^k \frac{(N_{ij} - e_{ij})^2}{e_{ij}}. \quad \text{Under the null hypothesis the distribution of } Y^2 \text{ is approximately that of } \chi^2 \text{ with } (h-1)(k-1) \text{ degrees of freedom.}$$

Two sided test only i.e.  $|Y^2|$ .

**Don't forget to pool  $e_{ij} < 5$**

### Exact test for 2x2 table

Discreet Distribution

Find smallest row/column as  $k$  then smallest other as  $n$  and  $x$  is the cell frequency.

$H_0$ : There is no association between attribute A and Attribute B

$H_1$ : Have to figure it out. (Is the value of  $x$  unusually to large or small to ascribe to chance.

One sided:

Two sided test  $\frac{\alpha}{2}$  reject  $H_0$  if  $x$  is a rare event.

## Correlation

$$R = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} \quad R = \frac{\sum X_i Y_i - \frac{(\sum x_i)(\sum Y_i)}{n}}{\sqrt{\left(\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right) \left(\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}\right)}}$$

In using R X and Y should follow conditions for bivariate normality:

- 1) Marginal normality
- 2) Linearity

### Testing for zero correlation $\rho = 0$

$H_0 : \rho = 0$  against the alternatives

$H_1 : \rho < 0$  or

$H_1 : \rho > 0$  or

$H_1 : \rho \neq 0$ .

$$T = \frac{\sqrt{n-2}R}{\sqrt{1-R^2}}$$

Student's t-distribution  $\sim T_{n-2}$

$$\sum (X_i - \bar{X})^2 = \sum X_i^2 - (\sum X_i)^2 / n$$

$$\sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - (\sum Y_i)^2 / n$$

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - (\sum X_i)(\sum Y_i) / n.$$

Correlation between X and Y is the same for any linear transformation of X or Y with positive coefficients.

### Fisher's Z-transformation for $\rho \neq 0$

#### Theorem 5.5

Let R be the sample correlation coefficient of a random sample from a bivariate normal distribution.

$$\text{Let } U = \frac{1}{2} \log_e \frac{1+R}{1-R} \text{ and } \eta = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho}.$$

Then, for large samples,  $Z = \sqrt{n-3}(U - \eta)$  is approximately a  $n(0; 1)$  variate.

$H_0 : \rho = \rho_0$  against

$H_1 : \rho > \rho_0$  or

$H_1 : \rho < \rho_0$  or

$H_1 : \rho \neq \rho_0$  we compute

### Confidence interval for $\rho$

Do algebraic transformation on  $1 - \alpha = P(a < Z < b)$  with  $Z$  =(fisher transform) and then use to get interval OR use table X inversely by linear interpolation.

$$\rho = \frac{e^\eta - e^{-\eta}}{e^\eta + e^{-\eta}} = \tanh(\eta)$$

## Equality of two R

$$H_0 : \rho_1 = \rho_2$$

If  $\rho_1 = \rho_2$ , that is  $\eta_1 = \eta_2$ , then  $Z = \frac{U_1 - U_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$  is approximately  $n(0; 1)$ .

## Study Unit 6

### Single sample

#### Hypothesis testing

We want to test the null hypothesis  $H_0: \sigma^2 = c$ .

##### (a) $\mu$ known

The procedure is based on the statistic  $U = \sum_{i=1}^n (X_i - \mu)^2 / c$  which, if  $H_0$  is true, is a  $\chi_n^2$  variate.

If  $\sum (X_i - \mu)^2$  is small, it is an indication that  $\sigma^2$  is small and vice versa. We reject  $H_0 : \sigma^2 = c$  against the alternatives

(i)  $H_1 : \sigma^2 \neq c$  if  $U < \chi_{1-\frac{1}{2}\alpha;n}^2$  or  $U > \chi_{\frac{1}{2}\alpha;n}^2$

(ii)  $H_1 : \sigma^2 < c$  if  $U < \chi_{1-\alpha;n}^2$

(iii)  $H_1 : \sigma^2 > c$  if  $U > \chi_{\alpha;n}^2$ .

##### (b) $\mu$ unknown

The procedure is based on the statistic  $U = \sum_{i=1}^n (X_i - \bar{X})^2 / c$  which, if  $H_0$  is true, is a  $\chi_{n-1}^2$  variate.

We reject  $H_0 : \sigma^2 = c$  against the alternatives

(i)  $H_1 : \sigma^2 \neq c$  if  $U < \chi_{1-\frac{1}{2}\alpha;n-1}^2$  or  $U > \chi_{\frac{1}{2}\alpha;n-1}^2$

(ii)  $H_1 : \sigma^2 < c$  if  $U < \chi_{1-\alpha;n-1}^2$

(iii)  $H_1 : \sigma^2 > c$  if  $U > \chi_{\alpha;n-1}^2$ .

### Two independent samples

$$F = \frac{(\chi_{n_1-1}^2) / (n_1 - 1)}{(\chi_{n_2-1}^2) / (n_2 - 1)}$$

yields test statistic

$$F = \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{S_1^2}{S_2^2} \sim F_{n_1-1; n_2-1}.$$

Hypotheses:

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = c$$

$$H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq c$$

$$H_1 : \frac{\sigma_1^2}{\sigma_2^2} < c$$

$$H_1 : \frac{\sigma_1^2}{\sigma_2^2} > c$$

$$F \sim F_{f,g} \text{ then } \frac{1}{F} \sim F_{g,f}$$

### Paired observations

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$T = \sqrt{n-2} \frac{U_{11} - U_{22}}{2\sqrt{U_{11}U_{22} - U_{12}^2}}$$

has a  $t_{n-2}$  distribution provided  $H_0 : \sigma_1^2 = \sigma_2^2$  is true.

$$U_{11} = \sum (X_{1j} - \bar{X}_1)^2$$

$$U_{22} = \sum (X_{2j} - \bar{X}_2)^2$$

$$U_{12} = \sum (X_{1j} - \bar{X}_1)(X_{2j} - \bar{X}_2)$$

Multiply out to get calculator formulas.

### More than two independent samples

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 \text{ against the alternative } H_1 : \sigma_p^2 \neq \sigma_q^2 \text{ for at least one } p \neq q.$$

Test Statistic

$$U = \max_i S_i^2 / \min_i S_i^2$$

$n - 1$  degrees of freedom,  $n$  = size of samples.

$k$  = number of sample variances.

## Study Unit 7

### One Sample Problem

(a)  $\bar{X}$  is a  $n(\mu; \sigma^2/n)$  variate, that is  $\sqrt{n}(\bar{X} - \mu) / \sigma$  is a  $n(0; 1)$  variate;

(b)  $(n-1)S^2/\sigma^2$  is a  $\chi_{n-1}^2$  variate; known variance

(c)  $\bar{X}$  and  $S^2$  are independent;

(d)  $T = \sqrt{n}(\bar{X} - \mu) / S$  is a  $t_{n-1}$  variate. unknown variance

95% Tolerance Interval: 95% confidence that at least 90% of the population data lie between  $x$  and  $y$ .

### Power of the test.

Given that  $Z_0$  is standard normal Solve for  $\bar{X}$  at significance level.

Then use  $\beta = P(\text{type II error}) = P(H_0 \text{ is not rejected} | H_1 \text{ is true})$

and standardise using do not reject signs and get probability

Then power of test is  $1 - \beta$



**Two-sample problem, independent samples. same variance****Theorem 7.2**

$$T = \frac{U}{\sqrt{\frac{W}{(n_1 + n_2 - 2)}}} \sim t_{n_1 + n_2 - 2}$$

$$\begin{aligned} \text{where } T &= \frac{[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)] / \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}{\sqrt{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2] / (n_1 + n_2 - 2)}} \\ &= \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \end{aligned}$$

$$\begin{aligned} \text{and } S_p^2 &= \frac{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]}{(n_1 + n_2 - 2)} \\ &= \frac{\left[ \sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2 \right]}{(n_1 + n_2 - 2)} \end{aligned}$$

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$

$$H_1 : \mu_1 - \mu_2 < 0$$

$$H_1 : \mu_1 - \mu_2 > 0$$

**Paired observations**

Same as one sample problem, by transforming by subtraction, then use  $T = \frac{\sqrt{n}(\bar{Y} - \mu)}{S_Y}$

**Independent samples with unequal variances**

We want to test  $H_0 : \mu_1 - \mu_2 = c$  (with  $c$  specified) against

$$H_1 : \mu_1 - \mu_2 \neq c \text{ or}$$

$$H_1 : \mu_1 - \mu_2 < c \text{ or}$$

$$H_1 : \mu_1 - \mu_2 > c.$$

We use the statistic

$$T = \frac{\bar{X}_1 - \bar{X}_2 - c}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

to get degrees of freedom, use formula and the interpolate table III.



$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}.$$

### More than two independent samples One-way ANOVA

$$SSE = \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 \quad MSE = \frac{SSE}{(kn - k)}$$

$$SSTr = n \sum_{i=1}^k (\bar{X}_i - \bar{X})^2 = \sigma^2 U \quad MSTR = \frac{n \sum_{i=1}^k (\bar{X}_i - \bar{X})^2}{k - 1} = \frac{\sigma^2 U}{(k - 1)}$$

#### Theorem 7.7

$$\text{Let } F = \frac{MSTR}{MSE} = \frac{\sigma^2 U / (k - 1)}{\sigma^2 V / (kn - k)} = \frac{n \sum_{i=1}^k (\bar{X}_i - \bar{X})^2 / (k - 1)}{\sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 / (kn - k)}.$$

Then  $F \sim F_{k-1;kn-k}$  if  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$  is true.

ANOVA table

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Treatments	0.675	3	0.225	5.625
Error	0.640	16	0.04	
Total	1.315	19		

### Multiple comparisons for ANOVA

$$T_{pq} = \frac{\bar{X}_p - \bar{X}_q}{S \sqrt{\frac{1}{n} + \frac{1}{n}}} = \frac{\sqrt{n}(\bar{X}_p - \bar{X}_q)}{\sqrt{2}S}$$

and reject  $H_0(p; q) : \mu_p = \mu_q$  in favour of

$H_1(p; q) : \mu_p \neq \mu_q$  if  $|T_{pq}|$  exceeds a critical value.

$$|T_{pq}| > \sqrt{(k-1) F_{\alpha; k-1; kn-k}}.$$

## Study Unit 8

The maximum likelihood estimators for  $\beta_0$ ,  $\beta_1$  and  $\sigma^2$  are

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum Y_i (X_i - \bar{X})}{\sum (X_i - \bar{X})^2}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2.$$

$$\hat{\beta}_1 = \frac{\sum (X - r\bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2}$$

**Theorem 8.1**

$\hat{\beta}_0$  and  $\hat{\beta}_1$  are jointly normally distributed with

$$E(\hat{\beta}_0) = \beta_0; \quad E(\hat{\beta}_1) = \beta_1;$$

$$Var(\hat{\beta}_0) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{X}^2}{d^2} \right); \quad Var(\hat{\beta}_1) = \frac{\sigma^2}{d^2};$$

$$Cov(\hat{\beta}_0; \hat{\beta}_1) = \frac{-\sigma^2 \bar{X}}{d^2} \quad \text{with } d^2 = \sum (X_i - \bar{X})^2.$$

**Theorem 8.2**

$$\frac{\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}{\sigma^2} \sim \chi_{n-2}^2$$

and is independent of  $\hat{\beta}_0$  and  $\hat{\beta}_1$ .

We call  $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$  the *predictions* and  $Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$  the *residuals*, that is the difference between the observations and predictions.

### Inference on the coefficients

**Theorem 8.2**

$$T_0 = \frac{\hat{\beta}_0 - \beta_0}{S \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{d^2}}} \text{ is a } t_{n-2} \text{ variate.}$$

**Theorem 8.3**

$$T_1 = \frac{\hat{\beta}_1 - \beta_1}{\frac{S}{d}} \text{ is a } t_{n-2} \text{ variate.}$$

Variance around regression line:

$$S^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2 - \hat{\beta}_1^2 \sum_{i=1}^n (X_i - \bar{X})^2}{(n-2)}.$$

Confidence interval for  $B_0$

$$\hat{\beta}_0 \pm t_{0.025; 13} SE(\hat{\beta}_0)$$

Confidence interval for  $B_1$

$$\hat{\beta}_1 \pm t_{0.025; 13} SE(\hat{\beta}_1)$$

### Inference on the regression line

$$\begin{aligned} E[\hat{Y}(X)] &= E(\hat{\beta}_0) + X E(\hat{\beta}_1) \\ &= \beta_0 + \beta_1 X \end{aligned}$$

$$\begin{aligned} Var[\hat{Y}(X)] &= Var(\hat{\beta}_0) + X^2 Var(\hat{\beta}_1) + 2X cov(\hat{\beta}_0; \hat{\beta}_1) \\ &= \sigma^2 \left[ \left( \frac{1}{n} + \frac{\bar{X}^2}{d^2} + \frac{X^2}{d^2} - \frac{2X\bar{X}}{d^2} \right) \right] \\ &= \sigma^2 \left[ \frac{1}{n} + \frac{(X - \bar{X})^2}{d^2} \right] \end{aligned}$$

$$(\hat{\beta}_0 + \hat{\beta}_1 X) \pm t_{\frac{\alpha}{2}; n-2} S \sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{d^2}}$$

### Relationship between test for correlation and regression

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

$$S^2 = \frac{1}{n-2} \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2.$$

$$Var(\sum b_i Y_i) = \sigma^2 \sum b_i^2$$

to standardise:  $\frac{X - E(X)}{\sqrt{Var(x)}}$