# SMI: Fucked shit we have to learn for exams

Andrew Martin

November 9, 2017

1. Estimation

   (a) Mean squared error

      i. If T is an estimator of $\theta$ then the **mean squared error** of T is:

      $$MSE_T(\theta) = E\left[(T-\theta)^2\right]$$

      ii. Inequality (i can't remember the name)
      For any $k > 0$

      $$P\left(|T-\theta| \geq k\sqrt{MSE}\right) \leq \frac{1}{k^2}$$

   (b) Bias

      i. Let T be an estimator for $\theta$ then the **bias** of T is defined by

      $$b_T(\theta) = E[T] - \theta$$

      ii. If $b_T(\theta) = 0$ for all $\theta$, then T is an **unbiased estimator** for $\theta$.

      iii. Theorem

      $$MSE_T(\theta) = var(T) + b_T(\theta)^2$$

   (c) BLUE (Best linear unbiased estimator)

      i. A linear estimator has form

      $$T = \sum_{i=1}^{n} a_i Y_i$$

      for constants $a_i$

      ii. Lemma $Y_i$ independent random vars with

      $$E[Y_i] = \mu_i \quad and \quad var(Y_i) = \sigma_i^2$$

      Let

      $$T = \sum_{i=1}^{n} a_i Y_i$$

      Then

      $$E[T] = \sum_{i=1}^{n} a_i \mu_i \quad and \quad var(T) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

iii. The Linear combination of normals is normal (if all the $Y_i$'s above were Normal, then T is normal.

iv. The BLUE for $\theta$ is the linear estimator with minimum variance

v. BLUE for $\mu$
$Y_i$ iid random vars with $E[Y_i] = \mu$ and $\text{var}(Y_i) = \sigma^2$ The BLUE for $\mu$ is

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

vi. Standard Error
If T is an unbiased estimator for $\theta$ then the standard deviation of the estimator is the **standard error**

$$SE(T) = \sqrt{var(T)}$$

(d) Confidence intervals

i. A random interval $(L, U)$ is a $100(1-\alpha)\%$ confidence interval for the parameter $\theta$ if it satisfies

$$P(L < \theta < U) = 1 - \alpha$$

ii. Theta is not random, L and U are.

iii. Normal CI for $\mu$. If $Y_i$ i i d $N(\mu, \sigma^2)$ with known $\sigma^2$ then

$$\left(\bar{y} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{y} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Is a $100(1-\alpha)\%$ CI for $\mu$ Where $P(Z > z_{\alpha/2}) = \alpha/2$

(e) Hypothesis Tests

i. General formulation

$$H_0 : \theta \in \Theta_0$$
$$H_a : \theta \notin \Theta_0$$

Where $\Theta$ is the parameter space

ii. Make sure to fucking say what all the params *are* or you're going to fail the course

iii. Test statistic

A. Normal:
Standard normal given $Y \sim N(\mu, \sigma^2)$

$$Z = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}}$$

iv. Retain/Reject When testing $H_0 : \mu = \mu_0$

A. P-value

$$\text{P-value} = P(|Z| \geq |z|)$$

(Area of the tails of the distribution)
*If the p-value is low, $H_0$ must go*

B. Confidence interval
If $(L, U)$ is a $100(1-\alpha)\%$ CI for $\theta$

- Reject $H_0 : \theta = \theta_0$ if $\theta_0 \notin (L, U)$
- Retain $H_0 : \theta = \theta_0$ if $\theta_0 \in (L, U)$

v. Type I and Type II errors

| | Retain $H_0$ | Reject $H_0$ |
|---|---|---|
| $H_0$ true | Correct conclusion | Type I error |
| $H_0$ false | Type II error | Correct conclusion |

vi. Significance level

$$\alpha = P(\text{reject } H_0 | H_0 \text{ true}) = P(\text{type I error})$$

vii. Power

$$\beta = P(\text{reject } H_0 | H_0 \text{ false}) = 1 - P(\text{type II error})$$

viii. one-sided test.
Testing:

$$H_0 : \mu \leq \mu_0$$

A. For normal:
Use the standard normal (Z) and reject $H_0$ if $z \geq z_\alpha$

B. type I error probability for onesided

$$P(\text{reject } H_0 | H_0 \text{ true}) \leq \alpha$$

2. Distributions from the Normal

(a) Standard Normal
$Y_i \sim N(\mu, \sigma^2)$ iid
Let

$$Z = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}$$

Then

$$Z \sim N(0, 1)$$

(b) Sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$$

$$E[S^2] = \sigma^2$$

(c) T-distribution
When using the sample variance

$$\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

(d) Chi-squared ($\chi^2$) distribution
$Z_i$ iid $N(0,1)$ random vars.

$$X = \sum_{i=1}^{p} Z_i^2 \sim \chi_p^2$$

Has p degrees of freedom

(e) More $\chi^2$
$Y_i$ iid $N(\mu, \sigma^2)$ then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

(f) More t-dist
$Z \sim N(0,1)$ and $X \sim \chi_p^2$ independently

$$T = \frac{Z}{\sqrt{X/p}} \sim t_p$$

Is t distributed with p degrees of freedom

(g) EVEN MORE t-dist
$Y_i$ iid $N(\mu, \sigma^2)$ then

$$\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

(h) F-dist
Let $W_1$ and $W_2$ be independent $\chi^2$ distributed random variables with $\nu_1$ and $\nu_2$ df respectively

$$F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F$$

With $\nu_1$ numerator df and $\nu_2$ denominator df.

(i) One-sample T-test $Y_i$ $i = 1, ..., n$ are iid $N(\mu, \sigma^2$ random variables with $\sigma^2$ known.

- the BLUE for $\mu$ is $\bar{Y}$
- the estimated standard error for $\bar{Y}$ is $S/\sqrt{n}$

i. Hypothesis test

$$H_0 : \mu = \mu_0,$$
$$H_a : \mu \neq \mu_0$$

ii. Test statistic

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{N}}$$

iii. Reject/retain

Reject iff

$$|t_{obs}| \geq t_{n-1,\alpha/2}$$

This has significance $\alpha$

iv. Confidence interval

$$\left(\bar{Y} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \bar{Y} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right)$$

is a $100(1-\alpha)\%$ confidence interval for $\mu$

(j) Inference for $\sigma^2$

$Y_i$ iid $N(\mu, \sigma^2)$, and $c_1, c_2$ be

$$P(c_1 < X < c_2) = 1 - \alpha$$

Where $X \sim \chi^2_{n-1}$. Then:

$$\left(\frac{(n-1)S^2}{c_2}, \frac{(n-1)S^2}{c_1}\right)$$

is a $100(1-\alpha)\%$ CI for $\sigma^2$

(NOTE THE ORDER OF THEM IS SWAPPED)

i. Choices for $c_1, c_2$
   - Symmetric: $P(X < c_1) = P(X > c_2) = \alpha/2$
   - Lower-bound: $c_1 = 0$ and $P(X > c_2) = \alpha$
   - Upper-bound: $P(X < c_1) = \alpha$ and $c_2 = \infty$

(k) Pivotal quantity

A random variable

$$H = H(Y_1, Y_2, ..., Y_n, \theta)$$

With a known distribution which doesn't depend on $\theta$ is called a **pivotal quantity**

(l) Two-sample t-test - pooled

i. Testing that the means are equal and assume the variances are the same

ii. Test the hypothesis

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

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

The test statistic is

$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Reject $H_0$ iff

$$|t_{obs}| \geq t_{n_1+n_2-2,\alpha/2}$$

Confidence interval:

$$\left(\bar{Y}_1 - \bar{Y}_2 - t_{n_1+n_2-2,\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{Y}_1 - \bar{Y}_2 + t_{n_1+n_2-2,\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

Is a $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$.

iii. Setup and derivation:
$Y_{ij}$ with $i = 1, 2$ $j = 1, 2, \ldots, n_i$ Such that:

$$Y_{ij} \sim N(\mu_i, \sigma^2)$$

iv. Sample means

$$\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \quad \text{for } i = 1, 2$$

Then

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right)$$

v. The pooled estimator

$$S_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Where

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_i\right)^2$$

Is an unbiased estimator for $\sigma^2$ and

$$\frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

Since $S_p^2$ is indep of $\bar{Y}_i$ it follows that

$$\frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

(m) Two-sample t-test - not pooled

i. Assume variances are not the same i.e.

ii. Setup and derivation:
$Y_{ij}$ with $i = 1, 2$ $j = 1, 2, \ldots, n_i$ Such that:

$$Y_{ij} \sim N(\mu_i, \sigma_i^2)$$

Where $\sigma_1^2 \neq \sigma_2^2$

$$\bar{Y_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \quad \text{for } i = 1, 2$$

then

$$\bar{Y_1} - \bar{Y_2} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

iii. Approximate the solution by choosing $t_k$ which approximates the true distribution of the test statistic (easy way and computer way)
Easy:

$$k = \min(n_1 - 1, n_2 - 1)$$

Computer:

$$k = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}$$

iv. Pooled vs not pooled
Use pooled if:

$$\frac{\max(s_1, s_2)}{\min(s_1, s_2)} < 2$$

3. Linear Models

(a) Data of form

$$(x_i, y_i), \quad i = 1, ..., n$$

(b) Linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

With

$$\epsilon_i \sim N(0, \sigma^2) \quad \text{independently for all } i$$

(c) Least squares estimation

$$\hat{\beta_0} = \bar{y} - \hat{\beta_1}\bar{x} \quad \text{and } \hat{\beta_1} = \frac{S_{xy}}{S_{xx}}$$

Where

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i \bar{y})$$

(d) Estimation of $\sigma^2$

$$s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2$$

(e) Expectations/variances:
Given $Y_i$ indep

$$E[Y_i] = \beta_0 + \beta_1 x_i \quad var(Y_i) = \sigma^2$$

then

$$E[\hat{\beta}_0] = \beta_0 \text{ and } E[\hat{\beta}_1] = \beta_1$$

$$var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) \text{ and } var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

$$cov(\hat{\beta}_0, \hat{\beta}_1 = -\frac{\bar{x}\sigma^2}{S_{xx}}$$

$$E[S_e^2] = \sigma^2$$

(f) Distribution of parameters

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}}{S_{xx}}\right)\right)$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

(g) Prediction (using the same setup):

$$E[\hat{\beta}_0 + \hat{\beta}_1 x_0] = \beta_0 + \beta_1 x_0$$

$$var[\hat{\beta}_0 + \hat{\beta}_1 x_0] = \sigma^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)$$

Then

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \sim N\left(\beta_0 + \beta_1 x_0, \sigma^2\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right)\right)$$

(h) Confidence interval
A $100(1-\alpha)\%$ CI for $E[\beta_0 + \beta_1 x_0]$ is:

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2} S_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

A $100(1-\alpha)\%$ CI for $Y_0$ (a new point)

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2} S_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

(i) Residuals:

$$\hat{e_i} = y_i - (\hat{\beta_0} + \hat{\beta_1} x_i)$$

$$\sum_{i=1}^{n} \hat{e_i} = 0$$

$$\sum_{i=1}^{n} \hat{e_i} x_i = 0$$

$$E[\hat{E_i}] = 0$$

$$var(\hat{E_i}) = \sigma^2 \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right)$$

(j) Standardised residuals:

$$\tilde{e_i} = \frac{y_i - (\hat{\beta_0} + \hat{\beta_1} x_i}{\sqrt{1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}}}$$

(k) Studentised residuals (i don't know what this means):

$$\tilde{e_i} = \frac{y_i - (\hat{\beta_0} + \hat{\beta_1} x_i}{s_e \sqrt{1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}}}$$

(l) Multiple linear regression:
consider $n$ subjects with $r$ predictors (n $y$'s and r $x$'s)
Then for $i = 1, 2, \ldots, n$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_r x_{ir} + \epsilon_i$$

Where

$$\epsilon_i \sim N(0, \sigma^2) iid$$

(m) Matrix form

$$Y = X\beta + \epsilon$$

I.e.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1r} \\ 1 & x_{21} & x_{22} & \ldots & x_{2r} \\ \vdots & & & & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nr} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

(n) Columns of X must be linearly independent (we all know what that means so i'm not writing it here)

(o) Lemma:
If $X_{nxp}$ is a matrix with linearly independent columns then $X^T X$ (which is $p$ x $p$) is invertible.

(p) LSE of $\beta$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

(q) Residual variance

$$S_e^2 = \frac{1}{n-p} ||Y - X\hat{\beta}||^2$$

Where $p = r + 1$ (the number of $\beta$'s)

(r) LSE dist inference

$$\lambda^T \hat{\beta} \sim N(\lambda^T \beta, \sigma^2 \lambda^T (X^T X)^- 1\lambda) \quad and \quad \frac{(n-p)S_e^2}{\sigma^2} \sim \chi^2_{n-p}$$

$$\frac{\lambda^T \hat{\beta} - \lambda^T \beta}{S_e \sqrt{\lambda^T (X^T X)^{-1} \lambda}} \sim t_{n-p}$$

(s) CI for $\lambda^T \beta$

$$\lambda^T \hat{\beta} \pm t_{n-p}(\alpha/2) s_e \sqrt{\lambda^T (X^T X)^{-1} \lambda}$$

(t) Hypothesis test

$$H_0 : \lambda^T \beta = \delta_0$$

$$t = \frac{\lambda^T \hat{\beta} - \delta_0}{s_e \sqrt{\lambda^T (X^T X)^{-1}}}$$

Reject if

$$|t| \geq t_{n-p}(\alpha/2)$$

(u) BLUE for MLR
$Y_i$ indep and $E(Y_i) = \eta_i$ and $var(Y_i) = \sigma^2$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \text{ and } \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}$$

And suppose $\eta = X\beta$ where $X$ is an $nxp$ matrix with lin indep columns.
If $a^T Y$ is an unbiased linear estimator for $\lambda^T \beta$ then:

$$var(a^T Y) \geq var(\lambda^T \hat{\beta})$$

With equality iff

$$a = X(X^T X)^{-1} \lambda$$

(v) Hypothesis Testing for several parameters.
Test the hypothesis:

$$H_0 : \beta_p = \beta_{p-1} = \ldots = \beta_{p-k+1} = 0$$

I.e. testing that the last $k$ components of $\beta$ are zero. Let $X_0$ contain the first $p-k$ columns of $X$ and let

$$\beta_0 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-k} \end{bmatrix}$$

Can test the hypothesis

$$H_0 : \eta = X_0 \beta_0$$

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T y \\ \hat{\eta} &= X\hat{\beta} \\ \hat{\beta_0} &= (X_0^T X_0)^{-1} X_0^T y \\ \hat{\eta_0} &= X_0 \hat{\beta_0} \end{aligned}$$

(w) Note

$$||y - X_0\hat{\beta_0}||^2 = ||y - X\hat{\beta}||^2 + ||X\hat{\beta} - X_0\hat{\beta_0}||^2$$

(x) If the null is not correct

$$E\left(\frac{1}{p-p_0}||X\hat{\beta} - X_0\hat{\beta_0}||^2\right) > \sigma^2$$

(y) Test $H_0$:

$$F = \frac{||X\hat{\beta} - X_0\hat{\beta_0}||^2/(p-p_0)}{||y - X\hat{\beta}||^2/(n-p)} \sim F_{p-p_0, n-p}$$

And reject if F is "Large"

4. Model Selection and ANOVA

(a) Forward selection

i. Begin with null model
ii. For each term not in the model calculate a P-value for including that term.
iii. If the smallest P-value is less than the threshold $\alpha = 0.05$, add it to the model

iv. Iterate 2-3 until no other terms are significant

(b) Backward selection

i. Begin with the fullest model (the most complicated to be considered)

ii. For each term not in the model calculate a P-value for removing that term.

iii. If the largest P-value is greater than the threshold $\alpha = 0.05$ remove it from the model

iv. Iterate 2-3 until the model only contains significant terms.

(c) Stepwise

i. Begin with the null model

ii. Perform one step of forward selection using a quite large $\alpha$ eg 0.2

iii. Perform a step of backward elimination with $\alpha = 0.05$

iv. Iterate 2-3 until no further changes occur or the algorithm cycles

(d) If an interaction term is included, all of its components must also be included

(e) Akaike Information Criterion (AIC) given k paramaters and $\log(\hat{L})$ is the log likelihood

$$AIC = 2k - 2\log(\hat{L})$$

Want to minimise AIC

(f) AICc for small sample sizes

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}$$

(g) Bayesian information criterion BIC

$$BIC = \log(n)k - 2\log(\hat{L})$$

(h) Cross-validation
Testing prediction by splitting the data into $k$ parts. Train on $k-1$ parts and test for the $kth$ part.

(i) Prediction error in cross validation

$$CV_{(K)} = \sum_{k=1}^{K} \frac{n_k}{n} MSE_k$$

$$MSE_k = \sum_{i \in C_k} \frac{(y_i - \hat{y}_i)^2}{n_k}$$

given n total observations and $n_k$ in the $kth$ group and $\hat{y}_i$ is the fitted value for the model with part $k$ removed

(j) Polynomial regression and two-sample pooled T-test

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_r x_i^2 + \epsilon_i$$

$$\epsilon_i \sim iidN(0, \sigma^2), \quad i = 1, 2, ..., n$$

Note this is linear in the coefficients $\beta_i$

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^r \\ 1 & x_2 & x_2^2 & \ldots & x_2^r \\ \vdots & & & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^r \end{bmatrix} and \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{bmatrix}$$

(k) If $x^r$ is in the model, must include all powers of x below that also

(l) This works the same way as normal multiple regression

(m) Interaction terms:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i$$

(n) X is what you would expect

(o) Two-sample pooled T-test
Given independent observations

$$Y_{ij} \sim N(\mu_i, \sigma^2) \text{ for } j = 1, 2, ..., n_i \quad i = 1, 2$$

(p) Set it as a MLR model

$$y = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \end{bmatrix}, X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}, \beta = \begin{bmatrix} \mu_1 \\ \mu_2 - \mu_1 \end{bmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \bar{Y_1} \\ \bar{Y_2} - \bar{Y_1} \end{bmatrix}$$

$$S_e^2 = \frac{1}{n-p} ||Y - X\hat{\beta}||^2 = \frac{(n_1 - 1)S_1^2 (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

(q) Hypothesis test

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

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

$$T = \frac{\bar{Y_2} - \bar{Y_1}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

(r) One-way ANOVA - THIS TOPIC IS LIKE ALL IN R.
Has 1 more paramater than value so set

$$\alpha_1 = 0$$

(s) Two-way ANOVA:
Have constraints of zero sums and include a reference category $\alpha_1 = \beta_1 = 0$

(t) ANCOVA

5. Likelihood Theory

(a) This part can be kinda fucked - be careful with order of paramaters

(b) Joint probability distributions
Consider indep random vars $Y_1, Y_2, \ldots, Y_n$ and let

$$f_i(y_i; \theta)$$

be the PDF if $Y_i$ is continuous and PMF if $Y_i$ is discrete. The joint PDF/PMF is

$$f(y; \theta) = \prod_{i=1}^{n} f_i(y_i; \theta)$$

(c) Likelihood function $L(\theta; y) = f(y; \theta)$

(d) Log likelihood $\ell(\theta; y) = \log L(\theta; y)$

(e) Score function $S(\theta; y) = \frac{\partial \ell}{\partial \theta}$

(f) MLE $\hat{\theta}$ is the $\theta$ which gives $S(\theta; y) = 0$

(g) Fisher information

$$I_\theta = E\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right]$$

$$I_\theta = -E\left[\frac{\partial^2 \ell}{\partial \theta^2}\right]$$

(h) Distribution of the score function:
Given $y_i$ indep observations with log-likelihood $\ell(\theta^*; y)$, where $\theta^*$ is the true value of $\theta$. Under certain regularity conditions on $f(y; \theta)$:

- $E(S(\theta^*; y)) = 0$
- $var(S(\theta^*; y)) = I_{\theta^*}$
- $\frac{S(\theta^*; Y)}{\sqrt{I_{\theta^*}}}$ converges to $N(0, 1)$ as $n \to \infty$
- $\hat{\theta} \sim N(\theta^*, I_{\theta^*}^{-1})$ Where $\hat{\theta}$ is the MLE for $\theta$

(i) Approximate CIs

$$\left(\hat{\theta} - z_{\alpha/2}\sqrt{I_{\hat{\theta}}^{-1}}, \hat{\theta} + z_{\alpha/2}\sqrt{I_{\hat{\theta}}^{-1}}\right)$$

(j) Wald test statistic:
Given iid $Y_i$

$$H_0 : \theta = \theta_0$$
$$Z = \frac{\hat{\theta} - \theta_0}{\sqrt{I_{\hat{\theta}}^{-1}}}$$

If the null hypothesis is true then $Z$ converges to $N(0,1)$ as $n \to \infty$.
So reject $H_0$ if $|Z| \geq z_{\alpha/2}$

(k) Score test statistic Given iid $Y_i$

$$H_0 : \theta = \theta_0$$
$$U = \frac{S(\theta_0; Y)}{\sqrt{I_{\theta_0}}}$$

If the null hypothesis is true then $U$ converges to $N(0,1)$ as $n \to \infty$.
So reject $H_0$ if $|U| \geq z_{\alpha/2}$

(l) Log-likelihood ratio test statistic Given iid $Y_i$

$$G^2 = -2\left(\ell(\theta_0; Y) - \ell(\hat{\theta}; Y)\right)$$

If $H_0$ is true then $G^2$ converges to $\chi_1^2$ as $n \to \infty$.
Reject if $g^2 \geq \chi_{1,\alpha}^2$

(m) Transformation of parameters:
Given an invertible, twice diffable function $\Phi$ can convert the paramater $\theta$ to $\phi = \Phi(\theta)$.

$$\ell_\phi(\phi; y) = \ell_\theta(\Phi^{-1}(\phi); y)$$

and

$$\ell_\theta(\theta; y) = \ell_\phi(\Phi(\theta); y)$$
$$\hat{\phi} = \Phi(\hat{\theta})$$

(n) Goodness of fit (doesn't get examined much).

(o) Multinomial dist $Y_i$ $i = 1, ...k$ follow multinomial distribution if their joint prob function is

$$p(y_1, ..., y_k) = \binom{n}{y_1, ..., y_k} \pi_1^{y_1} \ldots \pi_k^{y_k}$$

With

$$y_1, \ldots, y_k \geq 0 \quad y_1 + \ldots + y_k = n$$
$$\pi_1, \ldots, \pi_k \geq 0$$
$$\pi_1 + \ldots + \pi_k = 1$$

This gives

$$E(Y_i) = n\pi_i \quad var(Y_i) = n\pi_i(1 - \pi_i) \quad cov(Y_i, Y_j) = -n\pi_i\pi_j \quad i \neq j$$

(p) Goodness of fit test statistic

$$H_0 : \pi = \pi_0$$

$$X^2 = \sum_{i=1}^{k} \frac{(Y_i - n\pi_{0i})^2}{n\pi_{0i}}$$

If $H_0$ is true then for large $n$ the dist of $X^2$ is approximately $\chi^2_{k-1}$.
So reject if $x^2 \geq \chi^2_{k-1,\alpha}$
Given q paramaters have to be estimated to compute $\pi$ then reject if $x^2 \geq \chi^2_{k-q-1,\alpha}$

6. Bayesian Statistics

(a)

(b) Frequentist concept -
Make conclusions about an unknown parameter, $\theta$ given $y_i$ assumed to be observations from $f(y;\theta)$

(c) Prior
It is assumed that prior to observing data there is a prior distribution $p(\theta)$

(d) Posterior

$$p(\theta|y)$$

(e) Bayes' Thm

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta}$$

$$p(\theta|y) \propto p(\theta)p(y|\theta)$$

(f) Bayesian prediction
To predict $Y_0$ if $Y_0$ and $Y$ are conditionally independent given $\theta$

$$\begin{aligned} p(y_0|y) &= \int p(y_0, \theta|y)d\theta \\ &= \int p(y_0|\theta, y)p(\theta|y)d\theta \\ &= \int p(y_0|\theta)p(\theta|y0d\theta \end{aligned}$$

(g) Point estimation.
Supposing a point estimate of $\theta$ is needed

- Posterior mode : $\hat{\theta} = argmax_\theta p(\theta|y)$
- Posterior mean : $E(\theta|y) = \int \theta p(\theta|y) d\theta$
- Posterior median $\tilde{\theta}$ such that $P(\theta \leq \tilde{\theta}|y) = \frac{1}{2}$

(h) Bayesian credible interval
An interval $\ell$, $u$ is a $100(1-\alpha)\%$ BCI if:

$$P(\ell < \theta < u|y) = 1 - \alpha$$

(i) Conjugate priors
A family of priors $P = \{p(\theta)\}$ is conjugate to a family of likelihoods $\ell = \{p(x|\theta)\}$ if the posterior always satisfies $p(\theta|x) \in P$

(j) Trade-off between prior and data

- if n=0 the posterior reduces to the prior