S3 Notes

1 Estimation, confidence intervals and tests

1.1 Estimators

Bias is:

$$E(T) - \theta$$

To make the formula for S^2 in the formula book easier to use replace the top of the fraction with S_{xx} from the formula book

1.2 Standard error

Standard error is:

$$\frac{\sigma}{\sqrt{n}} \text{ or } \frac{s}{\sqrt{n}}$$

This can be remembered as the square root of the approximated variance given in the formula book

1.3 Central limit theorem

CLT states that sample means from a distribution where n is greater than 50 is:

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

1.4 Confidence intervals

To calculate the confidence interval use:

$$\overline{x} \pm z \times \frac{\sigma}{\sqrt{n}}$$

Where z is the value from the percentage points table that correlates to the probability in each of the tails on the normal distribution

2 Goodness of fit and Contingency tables

2.1 Goodness of fit method

Method for testing goodness of fit:

1. Determine which distribution would conceptually be most appropriate
2. Set significance level
3. Estimate parameters (if necessary) from observed data
4. Form hypotheses H_0 and H_1
5. Calculate expected frequencies
6. Combine expected frequencies so that none are < 5
7. Find degrees of freedom
8. Calculate critical value of χ^2 from the table
9. Calculate $\sum \frac{(O_i - E_i)^2}{E_i}$
10. See if the value is significant and draw conclusion

X^2 is distributed with a chi squared distribution χ^2_ν
Where ν = degrees of freedom

$$\text{The number of degrees of freedom} = \text{Number of classes (after combining)} - 1$$

2.2 Contingency tables

- We use this test to see if two factors are independent of each other
- We describe them by: Number of rows × Number of columns
- H_0 is that they are independent
- H_1 is that they are not independent
- $$\text{Expected values} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$
- $$\nu = \text{(Number of rows-1)(Number of columns-1)}$$

3 Combinations of normal distributions

3.1 Sums and Differences

$$\mu_{x+y} = \mu_x + \mu_y$$
$$\mu_{x-y} = \mu_x - \mu_y$$
$$\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y$$
$$\sigma^2_{x-y} = \sigma^2_x + \sigma^2_y$$

3.2 Multiples

$$\mu_{ax+by} = a\mu_x + b\mu_y$$
$$\mu_{ax-by} = a\mu_x - b\mu_y$$
$$\sigma^2_{ax+by} = a^2\sigma^2_x + b^2\sigma^2_x$$
$$\sigma^2_{ax-by} = a^2\sigma^2_x + b^2\sigma^2_x$$

3.3 Addition vs Multiplication

$$E(X_1 + X_2) = E(2X_1)$$

However

$$Var(X_1 + X_2) \neq Var(2X_1)$$

$$Var(X_1 + X_2) = \sigma_1^2 + \sigma_1^2 = 2\sigma_2^2$$
$$Var(2X_1) = 4\sigma_1^2$$

4 Regression and Correlation

Spearman's rank: The tendency for y to increase as x increases
PMCC:The closeness of the data to follow linear relationship

For example an x^3 graph would have a Spearman's rank of 1 as it always increases, however the PMCC would not be 1 as it does not follow a linear relationship.