ST2132 Cheatsheet

for finals, by ning

ANOVA

In this section, I denotes the treatments/groups, and J the measurements in each group. For two-factor, I, J denote the treatments/groups, and J the measurements within each I, J combination. Use ANOVA for comparing more than 2 groups.

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{\bar{Y}})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y})^2 + J\sum_{i=1}^{I}(\bar{Y}_i-\bar{\bar{Y}})^2$$

$$SS_{TOT} = SS_W + SS_B$$

$$SS_B/(I-1) \sim \chi^2_{I-1}$$

$$SS_W/[I(J-1)] \sim \chi^2_{I(J-1)}$$

$$\mathrm{E}[SS_W] = I(J-1)\sigma^2$$

$$\mathrm{E}[SS_B] = J\sum_{i=1}^{I}\alpha_i^2 + (I-1)\sigma^2$$

One-factor ANOVA, same-sized groups

The test statistic is

$$F = \frac{SS_B/(I-1)}{SS_W/[I(J-1)]} \sim F_{I-1,I(J-1)}$$

Reject H_0 if $F > F_{I-1,I(J-1)}(\alpha)$.

One-factor ANOVA, differently-sized groups

The test statistic follows a slightly different degree of freedoms $I-1 := df_1$ and $\sum_{i=1}^{n} J_i - I =: df_2$.

$$F \sim F_{df_1,df_2}$$

Two-factor ANOVA

There will be an additional sum of squares term for the interaction between groups. Its associated degree of freedom in as a chi-square distributed random variable, and within the final F distributed test statistic is $(I-1)(J-1)$. The sum of squared errors (within groups) has degree of freedom as chi-squared $IJ(K-1)$.

Post-ANOVA Tests

Turkey's correction and Bonferroni's correction reduces the probability of type I error in multiple tests after the ANOVA. The Kruskal-Wallis test, a generalisation of the Mann-Whitney test, is a nonparametric test which is particularly useful for small data sets.

Two-Sample Tests

In this section, $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ are each i.i.d. samples of X and Y respectively, unless otherwise stated. Define $D_i := X_i - Y_i$. The variances for X and Y are unknown.

Estimating the Equality of Variances

If $S_X \leq 2S_Y$ or $S_Y \leq 2S_X$, it is reasonable to assume that $\sigma_X = \sigma_Y$.

Normal, Unpaired, with Equal Variance

Calculate the pooled variance, s_p^2 as such.

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{m+n-2}$$

The test statistic t is

$$t = \frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{SE_{\bar{X}-\bar{Y}}}$$

which follows a t distribution with $m+n-2$ degrees of freedom.

Normal, Unpaired, with Unequal Variance

The variance of the sampling distribution $\mathrm{Var}(\bar{X}-\bar{Y})$ is simply

$$\mathrm{Var}(\bar{X}-\bar{Y}) = S_X^2/n + S_Y^2/m$$

The test statistic t is

$$t = \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{S_X^2/n + S_Y^2/m}}$$

which follows a t distribution with degrees of freedom df as

$$df = \frac{(S_X^2/n + S_Y^2/m)^2}{\frac{(S_X^2/n)^2}{n-1} + \frac{(S_Y^2/m)^2}{m-1}}$$

Normal, Paired

The variance of the sampling distribution $\mathrm{Var}(\bar{D})$ can be estimated simply by the unbiased sample variance of the series of D_i random variables. The test statistic t is

$$t = \frac{\bar{D}-\mu_D}{SE_{\bar{D}}}$$

Unpaired, Nonparametric

Rank the values of the samples X_i, Y_j from 1 to $n+m$. The null hypotheses is that X_i, Y_j are distributed identically, and therefore should rank 'evenly'. Then, the rank sum scores are defined as

$$R_X = \sum_{i=1}^{n}\mathrm{Rank}(X_i); \quad R_Y = \sum_{j=1}^{m}\mathrm{Rank}(Y_j)$$

Select the sample with the smaller size (w.r.t. n, m). Denote its rank sum score R, and define $R' := n(n+m+1) - R$. The Mann-Whitney test statistic is

$$R^* = \min(R, R')$$

H_0 is rejected for small R^*.

Paired, Parametric

Rank the *absolute* values of D_i from 1 to $n = m$. The null hypotheses is that D is distributed symmetrically about 0. Define

$$W_+ = \sum\{\mathrm{Rank}(D_i)|D_i > 0\}$$

$$W_- = \sum\{\mathrm{Rank}(D_i)|D_i < 0\}$$

Denote $W := \min(W_-, W_+)$. H_0 is rejected for small W.

Hypothesis Testing

In a hypothesis testing question, you must include (i) assumptions made, (ii) the null and alternate hypotheses, (iii) the test statistic and its distribution, (iv) the p-value, and (v) the conclusion.

Terminology

- The significance level (or size) α of a test is the probability of committing a *type I* error, or rejecting the null hypothesis, H_0 when it is true.
- The power $1-\beta$ of a test is the probability that H_0 is rejected when it is false.
- β denotes the probability of a *type II* error, or failing to reject H_0 when it is false.
- The α and power of a tests are mutual trade-offs.
- The set of values of a test statistic leading to rejection of H_0 is the rejection or critical region. Those leading to acceptance is the acceptance region.
- The p-value is the smallest significance level at which H_0 would be rejected.
- The null distribution is the probability distribution of the test statistic when H_0 is true.

Simple and Composite Hypotheses

A hypothesis that does not completely specify the probability distribution is called a composite hypothesis. Otherwise, it is a simple hypothesis. A hypothesis that is 'one-tailed' is called a 'one-sided' alternative.

Uniformly Most Powerful

If an alternative hypothesis H_1 is composite, a test that is most powerful for every simple alternative in H_1 is said to be uniformly most powerful. The test which is uniformly most powerful for a one-sided alternative is not for the two-sided.

Confidence Interval

Denote the acceptance region of the test as $A(\theta_0)$. Then, the set

$$C(X) = \{\theta | X \in A(\theta)\}$$

is a $100\%(1-\alpha)$ confidence region for θ. The CI contains all the values of θ for which the null hypothesis $H_0 : \theta = \theta_0$ is not rejected.

Neyman-Pearson Lemma

Suppose that H_0 and H_1 are simple hypotheses. Set the significance level of the test at α. Any other test for which the significance level is less than or equal to α has power less than or equal to that of the likelihood ratio test.

Generalised Likelihood Ratio Test (GLRT)

The generalised likelihood ratio test a non-optimal test used for situations of composite hypothesis where no optimal test exists. Denote the null and alternative hypotheses as $H_0 : \theta \in \omega_0$ and $H_1 : \theta \in \omega_1$ respectively, where ω_0, ω_1 are disjoint and subsets of Ω, the sample space. The generalised likelihood ratio test statistic is

$$\Lambda^* = \frac{\max_{\theta\in\omega_0} L(\theta)}{\max_{\theta\in\omega_1} L(\theta)}$$

For simplicity, we define Λ such that $\Lambda = \min(\Lambda^*, 1)$.

$$\Lambda = \frac{\max_{\theta\in\omega_0} L(\theta)}{\max_{\theta\in\Omega} L(\theta)}$$

Then, the generalised likelihood test rejects for $\Lambda \leq \lambda_0$, where $P(\Lambda \leq \lambda_0 | H_0) = \alpha$.

Distribution of $-2\log\Lambda$

For the GLRT, As the sample size $n \to \infty$, Under smoothness conditions on the pmfs or pdfs involved, the null distribution of $-2\log\Lambda$ tends to a chi-square distribution with degrees of freedom df as

$$df = \dim\Omega - \dim\omega_0$$

$\dim\Omega, \dim\omega_0$ are the number of free parameters under Ω and ω_0 respectively. Rejecting for small Λ is then also rejecting for large $-2\log\Lambda$. Special case: the one-tailed rejection region $-2\log\Lambda = n(\bar{X}-\mu_0)^2/\sigma^2 > \chi_1^2(\alpha)$ can be made two tailed $|\bar{X}-\mu_0| > (\sigma/\sqrt{n})z(\alpha/2)$ by definition of χ_1^2.

Likelihood Ratio Test (LRT)

In the case of the simple alternative hypothesis, simply define Λ directly.

$$\Lambda = \frac{L(\theta|H_0)}{L(\theta|H_1)}$$

Pearson Chi-square Test

The Pearson chi-square test is asymptotically equal to the GLRT. The test statistic for a multinomial distributed r.v. is

$$X^2 = \sum_{i=1}^{m}\frac{(O_i-E_i)^2}{E_i} = \sum_{i=1}^{m}\frac{(x_i-np_i(\hat{\theta}))^2}{np_i(\hat{\theta})}$$

Where $X^2 \sim \chi^2_{m-k-1}$, k is the number of values of the multinomial distribution.

Efficiency & Sufficiency

Mean Square Error (MSE)

The MSE is a common measure of accuracy of an estimator.

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}) &= \mathrm{E}[(\hat{\theta}-\theta_0)^2] \\ &= \mathrm{Var}(\hat{\theta}) + (\mathrm{E}[\hat{\theta}]-\theta_0)^2 \\ &= \mathrm{SE}^2 + \mathrm{bias}^2\end{aligned}$$

Efficiency

The efficiency of two estimators, $\hat{\theta}_0, \hat{\theta}_1$ is given as

$$\mathrm{eff}(\hat{\theta}_0, \hat{\theta}_1) := \mathrm{Var}(\hat{\theta}_1)/\mathrm{Var}(\hat{\theta}_0)$$

When any of the $\mathrm{Var}(\hat{\theta})$ is estimated via the asymptotic variance, the efficiency is called the asymptotic relative efficiency.

Cramér-Rao Inequality

Under smoothness assumptions of a $f(x|\theta)$ for a statistic $T := t(X_1, \cdots, X_n)$

$$\mathrm{Var}(T) \geq \frac{1}{nI(\theta)}$$

This gives the lower bound for the variance of any estimator of θ. An unbiased estimator whose variance achieves this lower bound is said to be efficient. The MLE is asymptotically efficient.

Sufficiency

A statistic $T(X_1, \cdots, X_n)$ is said to be sufficient for θ if the conditional distribution of $X_1, \cdots, X_n$ given $T = t$ does not depend on θ for any value of t. If T is sufficient for θ, the MLE for θ is a function only of T.

Factorization Theorem

The statistic $T(X_1, \cdots, X_n)$ is sufficient for a parameter θ iff the joint pdf factorises in the form

$$f(\vec{x}|\theta) = g(T(\vec{x}), \theta)h(\vec{X})$$

Exponential Family of Probability Distributions
1-parameter members of the exponential family have pdfs or pmfs in the form

$$f(x|\theta) = \begin{cases} \exp\{c(\theta)T(x) + d(\theta) + S(x)\}, & \text{if } x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

where the set A does not depend on θ.

Rao-Blackwell Theorem
Let $\hat{\theta}_0$ be an estimator for θ with finite second moment, T a sufficient statistic for θ, and $\hat{\theta}_1 = \mathrm{E}[\hat{\theta}_0|T]$.

$$\mathrm{E}[(\hat{\theta}_1 - \theta)^2] \leq \mathrm{E}[(\hat{\theta}_0 - \theta)^2]$$

$\hat{\theta}_1$ is an estimator of θ which is better than any estimator $\hat{\theta}_0$ since $\hat{\theta}_1 = \mathrm{E}[\hat{\theta}_0|T]$ which is a function of the sufficient statistic T.

Other Stuff

$$\frac{d}{dx} f(g(x)) = f'(g(x))g'(x)$$

$$\frac{d}{dx} f(x)g(x) = f(x)g'(x) + f'(x)g(x)$$

$$\frac{d}{dx} \frac{f(x)}{g(x)} = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$$

$$\frac{d}{dx} f(x)^{g(x)} = f(x)^{g(x)} \left(g'(x) \ln f(x) + g(x) \frac{f'(x)}{f(x)} \right)$$

$$\int_a^b u\, dv = uv\Big|_a^b - \int_a^b v\, du$$

$$\Gamma(z+1) = z\Gamma(z)$$
$$\Gamma(1) = 1$$
$$\Gamma(n) = 1 \cdot 2 \cdot \cdots \cdot (n-1) = (n-1)!$$

$$\mathrm{E}[X] = \sum_i x_i p(x_i)$$
$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx$$
$$\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}[Y|X]]$$
$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2]$$
$$\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2$$
$$\mathrm{Var}(a + bX) = b^2 Var(X)$$
$$\mathrm{Var}(Y) = \mathrm{Var}(\mathrm{E}[Y|X]) + \mathrm{E}[\mathrm{Var}(Y|X)]$$
$$\mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mu_X)(Y - \mu_Y)]$$
$$\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] \quad \text{if } X \perp\!\!\!\perp Y$$
$$\sum_{i=1}^n (X_i - \mu_0)^2 = \left[\sum_{i=1}^n (X_i - \bar{X})^2 \right] + n(\bar{X} - \mu_0)^2$$
$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$
$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\begin{aligned} &\mathrm{Cov}(aX + bY, cW + dV) \\ &= ac\,\mathrm{Cov}(X, W) + ad\,\mathrm{Cov}(X, V) + \\ &\quad bc\,\mathrm{Cov}(Y, W) + bd\,\mathrm{Cov}(Y, V) \end{aligned}$$

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{j=1}^n P(A|B_j)P(B_j)}$$

Information on Various Distributions

Binomial:

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n$$
$$\mathrm{E}[X] = np$$
$$\mathrm{Var}(X) = np(1-p)$$

Geometric:

$$p(k) = p(1-p)^{k-1}, \quad k = 1, \cdots$$
$$\mathrm{E}[X] = 1/p$$
$$\mathrm{Var}(X) = (1-p)/p^2$$

Negative binomial:

$$p(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \cdots$$
$$\mathrm{E}[X] = r/p$$
$$\mathrm{Var}(X) = [r(1-p)]/p^2$$

Poisson:

$$p(k) = (\lambda^k e^{-\lambda})/k!, \quad k = 0, 1, \cdots$$
$$\mathrm{E}[X] = \mathrm{Var}(X) = \lambda$$

Normal:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, \quad -\infty < x < \infty$$
$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}(X) = \sigma^2$$

Gamma:

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x \geq 0$$
$$\mathrm{E}[X] = \alpha/\lambda$$
$$\mathrm{Var}(X) = \alpha/\lambda^2$$

Chi-square:

$$Z \sim N(0, 1)$$
$$Z^2 \sim \chi_1^2 \sim \Gamma(1/2, 1/2)$$
$$Y_i \sim \chi_1^2;\ Y_1 + \cdots + Y_n \sim \chi_n^2, \quad \perp\!\!\!\perp Y_i$$

t:

$$\frac{Z}{\sqrt{U/n}} \sim t_n, \quad Z \sim N(0, 1);\ U \sim \chi_n^2$$
$$f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

F:

$$\frac{U/n}{V/m} \sim F_{n,m}, \quad U \sim \chi_n^2; V \sim \chi_m^2$$
$$f(x) = \frac{\Gamma[(n+m)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{n}{m}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} \left(1 + \frac{n}{m}x\right)^{-\frac{n+m}{2}}$$

$f(x)$ is over $x \geq 0$. If $T \sim t_n$ then $T^2 \sim F_{1,n}$.

Central Limit Theorem
For $S_n = \sum_{i=1}^n X_i$,

$$\lim_{x \to \infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) = \Phi(x)$$
$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Linear Functions of a Random Variable
Let $Y = g(X)$. To find $f_Y(y)$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= F_X(g^{-1}(y)) \\ f_Y(y) &= \frac{d}{dy} F_X(g^{-1}(y)) \\ &= \frac{dg^{-1}}{dy} f_X(g^{-1}(y)) \end{aligned}$$

Non-linear Functions of Random Variables
Let $Y = g(\vec{X})$, where $\vec{X} := (X_1, X_2, \cdots)$ with mean vector $\vec{\mu}$. Then, in order to find the mean and variance of Y, first take the Taylor expansion of $g(\vec{X})$,

$$\begin{aligned} Y &= g(\vec{X}) \\ &\approx g(\mu) + (X_1 - \mu_1)\frac{\partial g(\mu)}{\partial x_1} + (X_2 - \mu_2)\frac{\partial g(\mu)}{\partial x_2} + \cdots \end{aligned}$$

Then, $\mathrm{E}[Y] \approx g(\mu)$, and

$$Var(Y) \approx Var(g(\mu) + (X_1 - \mu_1) \cdots$$

Consider for example, $\vec{X} := (X_1, X_2)$. Then

$$\begin{aligned} Var(X) \approx\ & \sigma_{X_1}^2 \left(\frac{\partial g(\mu)}{\partial x_1}\right)^2 + \\ & \sigma_{X_2}^2 \left(\frac{\partial g(\mu)}{\partial x_2}\right)^2 + \\ & 2\sigma_{XY} \left(\frac{\partial g(\mu)}{\partial x_1}\right)\left(\frac{\partial g(\mu)}{\partial x_2}\right) \end{aligned}$$

Simple Random Sampling
Simple random sampling *without replacement* means that each sample is not independent of another. While the mean of the simple random sample is still unbiased, that is $\mathrm{E}[\bar{X}] = \mu$,

$$Cov(X_i, X_j) = -\sigma^2/(N-1)$$

for two different simple random samples, i.e. $i \neq j$. The variance of the sample mean then becomes

$$Var(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

The variance of the sample total is

$$Var(T) = N^2 \left(\frac{\sigma^2}{n}\right) \frac{N-n}{N-1}$$

σ is unknown and must be estimated.

$$s_{\bar{X}}^2 = \frac{s^2}{n}\left(1 - \frac{n}{N}\right)$$
$$s_T^2 = N^2 s_{\bar{X}}^2$$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is the unbiased sample variance.

Consistency
Let $\hat{\theta}_n$ be an estimate of a parameter θ_0 based on a sample of size n. $\hat{\theta}_n$ is said to be consistent in probability if $\hat{\theta}_n$ converges in probability to θ_0 as n approaches infinity. That is, for $\epsilon > 0$,

$$P(|\hat{\theta}_n - \theta_0| > \epsilon) \to 0 \text{ as } n \to \infty$$

Fisher Information

$$\begin{aligned} I(\theta) &= \mathrm{E}\left[\frac{\partial}{\partial\theta} \log f(X|\theta)\right]^2 \\ &= -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(X|\theta)\right] \end{aligned}$$

Large Sample Theory for MLE
Let $\hat{\theta}$ denote the MLE of θ_0. The probability distribution of

$$\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0)$$

tends to a standard normal distribution. Therefore, the asymptotic variance of the MLE is

$$\frac{1}{nI(\theta)} = -\frac{1}{\mathrm{E}[l''(\theta_0)]}$$

Approximate Confidence Intervals
Confidence intervals can be approximated through the large sample theory for MLE by taking $\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \to N(0, 1)$, as $n \to \infty$.

$$P\left(-z(\alpha/2) \leq \sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0) \leq z(\alpha/2)\right) \approx 1 - \alpha$$