Statistics Formula Sheet

William A. Bevington

Estimation and MLEs

An estimator T of θ is **unbiased** if:

$$\mathbb{E}(T) = \theta,$$

and **consistent** if:

$$Var(T) \to 0 \quad \text{and}$$
$$\mathbb{E}(T) \to \theta$$
$$\text{as} \quad n \to \infty$$

for example, $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$ is unbiased and consistent.

The **maximum likelihood estimation** $\hat{\alpha}$ is the solution to the equation

$$\frac{\partial}{\partial \alpha} L(x_1, \ldots, x_n; \theta) = \frac{\partial}{\partial \alpha} \prod_{i=1}^{n} f(x_i; \theta) = 0$$

where $f(x_i; \theta) = \mathbb{P}(X_i = x_i)$ and

Confidence Intervals:

These are the $100(1-\alpha)$ **confidence intervals** for various stuff. Define $X_1, \ldots, X_n \sim N(0,1)$, and remember that you **cannot** reject '$H_0 : p = p_0$' if $p_0 \in (a,b)$ where (a,b) is the confidence interval.

$$\mu;\ \sigma \text{ known}: \quad \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$
$$\mu;\ \sigma \text{ unknown}: \quad \bar{x} \pm t_{n-1;\alpha/2} \frac{\sigma}{\sqrt{n}}$$
$$\hat{\beta}: \quad \bar{\beta} \pm t_{n-2;\alpha/2} \times s.e.(\hat{\beta})$$
$$\hat{\alpha}: \quad \bar{\alpha} \pm t_{n-2;\alpha/2} \times s.e.(\hat{\alpha})$$
$$\hat{\mathbb{E}}(Y_0): \quad (\bar{\alpha} + \bar{\beta}x_0) \pm t_{n-2;\alpha/2} \sqrt{s^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}}\right)}$$
$$\hat{Y}_0: \quad (\bar{\alpha} + \bar{\beta}x_0) \pm t_{n-2;\alpha/2} \sqrt{s^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}}\right)}$$

Critical Regions

Remember that you **reject** H_0 if t, the observed test statistic, is **in** the critical region, ie $t = \frac{\bar{x}-65}{1.2}$ (pg.38 of the notes), reject H_0 if $t \leq -1.96 = -z_{\alpha/2}$ or if $t \geq 1.96 = z_{\alpha/2}$.

Type of test	Test Statistic	reject H_0 if:	For use when:
z-test	$\lvert T\rvert = \left\lvert \frac{X - \mu_0}{\sigma/\sqrt{n}} \right\rvert$	$\lvert t\rvert \geq z_{\alpha/2}$	σ^2 is known.
one sample t	$\lvert T\rvert = \left\lvert \frac{X - \mu_0}{S/\sqrt{n}} \right\rvert$	$\lvert t\rvert \geq t_{n-1;\alpha/2}$	σ^2 is unknown.
paired t	$\lvert T\rvert = \left\lvert \frac{D}{S/\sqrt{n}} \right\rvert$	$\lvert t\rvert \geq t_{n-1;\alpha/2}$	data is paired as (X_i, Y_i).
two sample t	$\lvert T\rvert = \left\lvert \frac{\bar{X} - \bar{Y}}{S_p/\sqrt{\frac{1}{m} + \frac{1}{m}}} \right\rvert$	$\lvert t\rvert \geq t_{m+n-2;\alpha/2}$	$\bar{X} \sim N(\mu_X, \frac{\sigma^2}{n})$ and $\bar{Y} \sim N(\mu_Y, \frac{\sigma^2}{m})$
F-test	$\lvert T\rvert = \left\lvert \frac{rss_0 - rss_1/m}{rss_1/n} \right\rvert$	$\lvert t\rvert \geq F_{m,n;\alpha/2}$	ANOVA Tests

and p-values are found using: p-value $= \mathbb{P}(|T| \geq t)$

Various Distributional Stuff

Chi-Squared

The chi-squared χ^2 distribution is defined by

$$\sum_{i=1}^{n} Z^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu_i}{\sigma}\right)^2 \sim \chi_n^2,$$

where each $Y_i \sim N(\mu_i, \sigma)$, and has the that

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

t-distribution

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} = \frac{Z}{\sqrt{Y/(n-1)}} \sim t_{n-1}$$

where $Z = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$ and $Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ are *independent* variables.

F-distribution

Let U and V be rvs with $U \sim \chi_{m-1}^2$ and $V \sim \chi_{n-1}^2$, then

$$\frac{U/m}{V/n} = \frac{(rss_0 - rss_1)/m}{rss_1/n} \sim F_{m,n}$$

where $rss_0 = SS_{Tot}$ and $rss_1 = SS_W$ are *residual sums of squares* (and $SS_B = rss_0 - rss_1$). Also we have that

$$\frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

And lastly, if $X \sim t_n$ then $X^2 \sim F_{1,n}$.

Assumptions of linear models

When forming linear models, the assumptions are that **the observations:**

- are Normally distributed
- are Independent
- have constant variance
- are linearly regressed

ANOVA

When conducting ANOVA tests, the *assumptions* are that **the observations:**

- are Normally distributed
- are Independent
- have constant variance
- have expected values dependent only on their groups

What ANOVA test do I use? (WANOVATDIU?)

One-Way if testing blocks and treatments that have single entries, and
Two-Way if you're concerned with seeing if the blocks and treatments have individual effects on a variable.

LSD

Remember to ONLY USE LSD's IF YOU REJECT H_0!

One-Way:

$$L.S.D. = t_{n-k;\alpha/2}\sqrt{s^2\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

where n_i and n_j are sample sizes for samples i and j.

Two-Way: No Replications

For Blocks:

$$L.S.D. = t_{(b-1)(k-1);\alpha/2}\sqrt{2\frac{s^2}{k}}$$

For Treatments:

$$L.S.D. = t_{(b-1)(k-1);\alpha/2}\sqrt{2\frac{s^2}{b}}$$

Two-Way: With Replications

For Blocks:

$$L.S.D. = t_{rbk-b-k+1;\alpha/2}\sqrt{2\frac{s^2}{rk}}$$

For Treatments:

$$L.S.D. = t_{rbk-b-k+1;\alpha/2}\sqrt{2\frac{s^2}{rb}}$$

where r is the number of replications per group.