Inferential Statistics Formula Review

I. Large sample hypothesis testing ($n \geq 30$)

A. One sample mean

1. One-tail testing determines if a mean is different than a given value in a particular direction. $H_0: \mu \geq x$ and $H_1: \mu < x$; $H_0: \mu \leq x$ and $H_1: \mu > x$
2. Two-tail testing determines if a mean is different than a given value in either direction. Divide α by 2. $H_0: \mu = x$ and $H_1: \mu \neq x$
3. The test statistic is $\bar{x}$.

X is the hypothesized population mean.

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

B. Two sample means

1. One-tail testing determines if one mean is larger or smaller than another. $H_0: \mu_1 \geq \mu_2$ and $H_1: \mu_1 < \mu_2$; $H_0: \mu_1 \leq \mu_2$ and $H_1: \mu_1 > \mu_2$
2. Two-tail testing determines if 2 means are equal. Divide α by 2. $H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 \neq \mu_2$
3. The test statistic is $\bar{x}$.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

C. One sample proportion

1. One-tail testing determines if a proportion is different than a given value in a particular direction. $H_0: p \geq x$ and $H_1: p < x$; $H_0: p \leq x$ and $H_1: p > x$
2. Two-tail testing determines if a proportion is different than a given value in either direction. Divide α by 2. $H_0: p = x$ and $H_1: p \neq x$
3. The test statistic is $\bar{p}$.

P is the hypothesized population proportion.

$$z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Gjjj

D. Two sample proportions

1. One-tail testing determines if one proportion is larger or smaller than another.
2. Two-tail testing determines if 2 proportions are equal. Divide α by 2.
3. The test statistic is $\bar{p}$.

$$\bar{p} = \frac{x}{n} \qquad z = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{\bar{p}w(1-\bar{p}w)}{n_1} + \frac{\bar{p}w(1-\bar{p}w)}{n_2}}} \quad \text{and} \quad \bar{p}_w = \frac{\text{Total successes}}{\text{Total sampled}} = \frac{x_1 + x_2}{n_1 + n_2}$$

II. Small sample hypothesis testing ($n < 30$)

A. One sample mean

1. One-tail testing determines if a mean is different from a given value in a particular direction.
2. Two-tail testing determines if a mean is different from a given value in either direction. Dividing α by 2.
3. The test statistic is $\bar{x}$.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad \text{and} \quad df = n - 1$$

B. Two sample means from independent populations

1. One-tail testing determines if one mean is larger or smaller than another.
2. Two-tail testing determines if 2 means are equal. Divide α by 2.
3. The test statistic is $\bar{x}$.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_w^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{and} \quad s_w^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} \quad \text{and } df = n_1 + n_2 - 2$$

C. Two sample means from dependent populations (paired difference test)
   1. One- and two-tail problems may be analyzed.
   2. The test statistic is $\bar{d}$.

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} \text{ and } s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} \text{ and } \bar{d} = \frac{\sum d}{n} \text{ and df = n - 1}$$

   3. $H_0 : \mu_d \geq 0$ and $H_1 : \mu_d < 0$ Note: μ_d is negative when H_1 involves testing for an increase.

III. Statistical quality control A. The $\bar{x}$ chart B. The R chart C. The p chart

IV. Analysis of variance
A. Testing 2 sample variances from normal populations
   1. One- and two-tail problems may be analyzed.
   2. The test statistic is F. $F = \frac{s_1^2}{s_2^2}$ df = n - 1 for both the numerator and the denominator
      Two-tail test requires dividing the level of significance by 2.

B. Analyzing 3 or more sample means from normally distributed populations (ANOVA)
   1. Equality of the means will be tested. $H_0 : \mu_1 = \mu_2 = \mu_3$ and $H_1 : \mu_1 \neq \mu_2 \neq \mu_3$
   2. The test statistic is F. $F = \frac{MS_T}{MS_E}$
   3. This is a one-tail test.

C. Two-factor variance analysis
   1. Equality of 3 or more means will be tested for both a treatment variable and a blocking variable.
   2. The test statistic is F. $F = \frac{MS_T}{MS_E}$ and $F = \frac{MS_B}{MS_E}$
   3. This is a one-tail test.

D. Comparing three or more treatment means to each other
   1. Having rejected the null hypothesis when comparing the means of three or more populations, treatment means can then be compared (2 at a time) to determine individual differences.
   2. The test statistic is the range for the difference between the treatments. If the range includes 0, conclude there is not a difference. $(\bar{x}_3 - \bar{x}_1) \pm t\sqrt{MS_E(\frac{1}{n_1} + \frac{1}{n_2})}$
   3. This is a two-tail test.

V. Nonparametric hypothesis testing
A. **Goodness of fit tests** for expected frequency of one categorical variable
   1. Do expected frequencies (equal or proportional) match the observed frequency?
   2. The test statistic is chi-square.

$$\chi^2 = \sum\left[\frac{(f_o - f_e)^2}{f_e}\right] \text{ and } f_e \geq 5 \text{ and df = k - 1}$$

B. Measuring independence of two categorical variables with a **contingency table test**
   1. Are two variables dependent?
   2. The test statistic is chi-square. $\chi^2 = \sum\left[\frac{(f_o - f_e)^2}{f_e}\right]$ and $f_e = \frac{f_r \times f_c}{n}$ $f_o \geq 5$, and df = (r - 1)(c - 1)

C. The **run test** for determining randomness based upon order of occurrence

$Z = \frac{r - \mu_r}{\sigma_r}$ where r is the number of runs, $\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1$ and $\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_1+n_2-1)}}$

D. One- and two-tail testing of one sample median using a **sign test**.

E. One- and two-tail testing of 2 medians from independent populations using the **Mann-Whitney test**.

$$z = \frac{U - \mu_U}{\sigma_U} \text{ where } U_1 = n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1 \text{ and } \mu_U = \frac{n_1n_2}{2} \text{ and } \sigma_U = \sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}$$

F. One- and two-tail testing of 2 medians from dependent populations using the **paired difference sign test**.

G. The **Kruskal-Wallis** test for the equality of 3 or more independent sample medians

$$H = \frac{12}{N(N+1)}\left[\frac{(\sum R_1)^2}{n_1} + \frac{(\sum R_2)^2}{n_2} + \ldots + \frac{(\sum R_k)^2}{n_k}\right] - 3(N+1)$$