FORMULAS

Sample statistics

Sample mean: $\bar{x}=\frac{\sum x_i}{n}$, Sample standard deviation (definition) $s=\sqrt{\frac{\sum (x_i-\bar{x})^2}{n-1}}$

Computational formula $s=\sqrt{\frac{\sum x_i^2-\frac{(\sum x_i)^2}{n}}{n-1}}$

Population parameters:

Standard score or z-score $z=\frac{x-\mu}{\sigma}$ If $x\sim N(\mu,\sigma)$ then $z\sim N(0,1)$

Sampling Distribution of $\bar{x}$

$\mu_{\bar{x}}=\mu$, $\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}$, Standardized version of $\bar{x}$: $z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Studentized version of $\bar{x}$: $t=\frac{\bar{x}-\mu}{s/\sqrt{n}}$

Confidence Intervals for μ

Confidence level C= $(1-\alpha)*100\%$

Z-interval: $\bar{x}\pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ Margin of error: $E=z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, Sample size estimation: $n=\left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$

t-interval: $\bar{x}\pm t_{\alpha/2}\frac{s}{\sqrt{n}}$, df=*n-1*

Hypothesis test for one Population Mean

Z-test (σ known): $H_0:\ \mu=\mu_0$ vs $H_a:\mu\neq\mu_0$ or $H_a:\mu>\mu_0$ or $H_a:\mu<\mu_0$

test statistics: $z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

t-test (σ unknown): $H_0:\ \mu=\mu_0$ test statistics: $t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$, df=*n-1*

Inferences for Two Population Means

Pooled sample standard deviation: $s_p=\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$

Pooled t-test: ($\sigma_1=\sigma_2$ unknown, independent samples) $H_0:\ \mu_1=\mu_2$ vs
$H_a:\mu_1\neq\mu_2$ or $H_a:\mu_1>\mu_2$ or $H_a:\mu_1<\mu_2$

Test statistics: $t=\frac{\bar{x}_1-\bar{x}_2}{s_p\sqrt{1/n_1+1/n_2}}$, $df=n_1+n_2-2$

Pooled t-interval for $\mu_1-\mu_2$: $(\bar{x}_1-\bar{x}_2)\pm t_{\alpha/2}\, s_p\sqrt{1/n_1+1/n_2}$

Confidence level C= $(1-\alpha)*100\%$

Paired t-test: (matched samples) $H_0:\ \mu_1=\mu_2$, $t=\dfrac{\bar{d}}{s_d/\sqrt{n}}$, df=*n-1*

Paired t-interval: $\bar{d}\pm t_{\alpha/2}\dfrac{s_d}{\sqrt{n}}$ Margin of error: $E=t_{\alpha/2}\dfrac{s_d}{\sqrt{n}}$,

Confidence level C= $(1-\alpha)*100\%$

Inferences for Population Proportions

One Population:

Sample proportion: $\hat{p}=\dfrac{x}{n}$, where x=number of members in a sample with specified attribute

One sample z-interval for *p*: $\hat{p}\pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, *x* and *p-x* both 5 or greater

Sample size for estimating p: $n=.25\left(\dfrac{z_{\alpha/2}}{E}\right)^2$ or $n=\hat{p}_g(1-\hat{p}_g)\left(\dfrac{z_{\alpha/2}}{E}\right)^2$

One sample z test: $H_0:\ p=p_0$ vs $H_a: p\neq p_0$ or $H_a: p>p_0$ or $H_a: p<p_0$

test statistics: $z=\dfrac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$, $np_0,\ n(1-p_0)$ both 5 or greater

Two Populations: independent samples, $x_1,\ n_1-x_1,\ x_2,\ n_2-x_2$ all 5 or greater

Pooled sample proportion: $\hat{p}_p=\dfrac{x_1+x_2}{n_1+n_2}$

Two samples z test: $H_0:\ p_1=p_2$ vs $H_a: p_1\neq p_2$ or $H_a: p_1>p_2$ or $H_a: p_1<p_2$

test statistics: $z=\dfrac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}_p(1-\hat{p}_p)(1/n_1+1/n_2)}}$

Two samples z-interval for p_1-p_2 : $(\hat{p}_1-\hat{p}_2)\pm z_{\alpha/2}\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1+\hat{p}_2(1-\hat{p}_2)/n_2}$

Chi-square Tests

Goodness-of-fit test: $H_0:$ The variable has specified distribution

$E=np$, df=*k-1, k*=number of categories

Test of independence: $H_0:$ Two variables are not associated

$E=\dfrac{R*C}{n}$, R=row total, C=column total , df=*(r-1)(c-1)*, *r*=#of rows, *c*=#of columns

Test statistics for each test : $\chi^2=\sum\dfrac{(O-E)^2}{E}$

For each test assume that all *E* are 1 or greater, and at most 20% of all *E* are < 5