A Statistics Summary-sheet

Is n is large, say over 30?

	Sampling Conditions	Confidence Interval	Test Statistic
Yes	σ^2 is known $\Rightarrow$ $\overline{X} \sim$ N (μ, σ^2/n)	$\overline{X} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$	$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$
Yes	σ^2 is unknown $\Rightarrow$ $\overline{X}$ ~ N (μ, σ^2/n)	$\overline{X} \pm Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$	$Z = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$
		$\overline{p} \pm Z_{\alpha/2}\left(\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right)$	$Z = \frac{\overline{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
No	X ~ N (μ, σ^2) and σ^2 is known $\Rightarrow$ $\overline{X}$ ~ N (μ, σ^2/n)	$\overline{X} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$	$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$
No	X ~ N (μ, σ^2) and σ^2 is unknown $\Rightarrow$ $\overline{X}$ ~ t_{n-1} (μ, σ^2/n)	$\overline{X} \pm t_{n-1,\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$	$t_{n-1} = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$

If n is not large, say over 30 and X is not ~ N (μ, σ^2), cannot proceed with parametric statistics.

Formulas, Distributions, and Concepts

Counting and Probabilities

$$_{n}P_{x} = \frac{n!}{(n-x)!}$$ Permutations

$$_{n}C_{x} = \frac{n!}{x!(n-x)!}$$ Combinations

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$ Conditional Probability

$$P(A \cap B) = P(A \mid B)P(B)$$ Probability of an Intersection

Discrete Probability Distributions

$$P_{x}(x) = \frac{n!}{x!(n-x)!} p^{x}(1-p)^{n-x}$$ Binomial Probability

$$P_{x}(x) = \frac{e^{-\mu}\mu^{x}}{x!}$$ Poisson Probability

Continuous Probability Distributions

Random Variable ~ Distribution (mean, variance)

Standard Normal $Z \sim N(0,1)$

Normal $X \sim N(\mu,\sigma^2)$

Binomial $X \sim \text{Binomial}\ [np, np(1-p)]$

Sample Mean $$\bar{X} \sim N\left(\mu\ , \frac{\sigma^2}{n}\right)$$

By CLT, if n ≥30, $$\bar{X} \sim N\left(\mu\ , \frac{\sigma^2}{n}\right)$$

Sample Proportion $$\bar{p} \sim \left(p, \frac{p(1-p)}{n}\right)$$

Confidence Intervals (Interval Estimation)

$$\bar{X} \pm z_{(\alpha/2)} \frac{\sigma}{\sqrt{\text{n}}}$$ If population is normal and population variance is known.

$$\bar{X} \pm z_{(\alpha/2)} \frac{s}{\sqrt{n}}$$ If population variance is unknown and n ≥30.

$$\bar{X} \pm t_{(n-1,\alpha/2)} \frac{s}{\sqrt{n}}$$ If population is normal, population variance is unknown.

$$\bar{p} \pm z_{(\alpha/2)} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$ If n ≥30.

$$(\bar{X} - \bar{Y}) \pm z_{(\alpha/2)} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$ If independent samples and either population variance known, or n ≥30 in which case, substitute sample variance for population variance.

$$(\bar{X} - \bar{Y}) \pm t_{(n_X + n_Y - 2,\alpha/2)} \sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$ where

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$ If independent samples, population variances unknown, but statistically equal.

Estimating Sample Size

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{E^2}$$ For estimation and CI for the population mean, normal population, σ^2 known, or estimated by a pilot run. E = absolute error.

Hypothesis Testing

1. Set up the **appropriate** null which must be in equality form, always and alternative hypotheses.
2. Define the rejection area. Take care as to whether the test is one-tailed or two-tailed. Look to the alternative hypothesis to determine this.
3. Calculate the test statistic.
4. State Decision.
5. Interpret your conclusion.

Hypothesis Testing (test statistics and their distributions under the null)

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim z_\alpha$$ When population variance known, or if n ≥30, substitute *s* for σ.

$$\frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1,\alpha}$$ When If population is normal, population variance unknown.

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)_0}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim \mathrm{t}_{n_1+n_2-2,\alpha}$$ where $s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ Difference in means, independent samples, population variances unknown, but statistically almost equal.

$$\frac{\bar{\mathrm{p}} - \mathrm{p}_0}{\sqrt{p_0(1-p_0)/\mathrm{n}}} \sim \mathrm{Z}_\alpha$$ Population Proportion, n $\geq$30

Formulas for ANOVA

$$\bar{X}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$$

$$s_j^2 = \frac{\sum_{i=1}^{n_j}\left(x_{ij}-\bar{X}_j\right)^2}{n_j-1}$$

$$\bar{\bar{X}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n_T}$$ Overall Sample Mean

or if the treatment sample sizes are all equal, $\bar{\bar{X}} = \frac{\sum_{j=1}^{k} \bar{X}_j}{k}$

$$MSTR = \frac{SSTR}{k-1} \qquad SSTR = \sum_{j=1}^{k} n_j \left(\bar{X}_j - \bar{\bar{X}}\right)^2$$

$$MSE = \frac{SSE}{n_T - k} \qquad SSE = \sum_{j=1}^{k} (n_j - 1) s_j^2$$

$$F = \frac{MSTR}{MSE} \sim F_{k-1, n_T - k}$$

<u>Terms and Concepts</u>

Central Limit Theorem: If the sample size n is large, say n $\geq$30 no matter what the population distribution is, the sampling distribution of the sample mean tends towards the normal as n gets large.