CONFIDENCE INTERVALS AND HYPOTHESIS TESTS

Situation	**Test Statistic**	**Distribution**	$100(1-\alpha)\%$ **CI**
Large sample, known σ	$Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	Approx Std Normal	$\overline{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
Normally distributed population, known σ	$Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	Std Normal	$\overline{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
Large sample, unknown σ	$Z = \dfrac{\overline{X}-\mu}{S/\sqrt{n}}$	Approx Std Normal	$\overline{x} \pm z_{\alpha/2}\dfrac{s}{\sqrt{n}}$
Normally distributed population, unknown σ	$T = \dfrac{\overline{X}-\mu}{S/\sqrt{n}}$	t-dist $n-1$ df	$\overline{x} \pm t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}}$
Large independent samples from 2 pops, known σ_1 and σ_2	$Z = \dfrac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	Approx Std Normal	$\overline{x}_1-\overline{x}_2 \pm z_{\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$
Independent samples from 2 pops, unknown variances. Large samples or normal pops.	$T = \dfrac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$	t-dist ν df	$\overline{x}_1-\overline{x}_2 \pm t_{\alpha/2,\nu}\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}$
Independent samples from 2 pops, unknown variances, $\sigma_1=\sigma_2$. Large samples or normal pops.	$T = \dfrac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{S_P^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$	t-dist, n_1+n_2-2 df	$\overline{x}_1-\overline{x}_2 \pm t_{\alpha/2,n_1+n_2-2}\sqrt{s_P^2\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}$

I. Formula for ν. Round down to get an integer value.

$$\nu \approx \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_2}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

II. Formula for the pooled estimator of a common variance

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

Date: April 5, 2024.

1. Confidence intervals

1. This problem deals with the US Census bureau's 2017 American Community Survey (ACS). The survey reports mean income along with a standard error $\frac{s}{\sqrt{n}}$.

a) The survey included 19,427 households in the Pacific West; these households had a mean income of \$101,716 with a standard error of \$1,584. Calculate a 99% confidence interval for the true mean income of a household in the Pacific West.

b) The survey also included 9,669 Mountain West households; these households had a mean income of \$88,739 with a standard error of \$1,746. Calculate a 99% confidence interval for the difference between the mean household incomes of these regions.

2. In a random sample of 16 games in 2016, the Gonzaga men's basketball team had an average score of $\overline{x} = 81.8750$ with a sample standard deviation of $s = 10.7881$. Calculate a 95% confidence lower bound for the mean score (assuming that scores are normally distributed).

3. In a random sample of 9 games in 2019, the men's basketball team had a mean score of 88.4444 with a sample standard deviation of 8.7050. Calculate a 95% confidence upper bound for the difference between the mean scores in 2016 and 2019. (Assume that scores for both years are normally distributed with the same variance).

4. Suppose that we want to predict Gonzaga's score in the next game (instead of producing confidence intervals for mean scores). This means that we should use a $100(1-\alpha)\%$ **prediction interval**:

$$\boxed{\overline{x} \pm t_{\alpha/2,n-1}\sqrt{s^2\left(1+\frac{1}{n}\right)}}$$

Use this formula to calculate a 95% prediction interval for the next score in 2019.

When estimating a population proportion θ we often use the sample proportion p. Now we can use $Z \approx \dfrac{P-\theta}{\sqrt{\frac{P(1-P)}{n}}}$, which is approximately standard normal as long as both $np \geq 8$ and $n(1-p) \geq 8$. This gives us the approximate $100(1-\alpha)\%$ CI for θ: $\boxed{p \pm z_{\alpha/2}\sqrt{\dfrac{p(1-p)}{n}}}$

5. As 2014 Pew Center study on religion in America surveyed a total of 35,071 Americans. Of those, 714 lived in Washington state. Of those living in Washington, 121 said they were Catholic.

a) Calculate a 98% confidence interval for the proportion of Americans who live in Washington.

b) Calculate a 98% confidence upper bound for the proportion of Washingtonians who are Catholic.

2. Hypothesis tests

Start with a **null hypothesis** H_0, which we assume to be true until we have evidence to the contrary. Note that H_0 needs to be based on prior knowledge, not the data we collect. Exactly what constitutes contrary evidence is determined by our choice of **alternative hypothesis** H_1. The null hypothesis should specify the distribution value of a population parameter, which allows us to calculate a **test statistic**. We collect data, calculate the observed value of a test statistic, then calculate the **P-value** (or observed significance level) of our data. There are 2 possible outcomes for a hypothesis test:

1. "Reject H_0 in favor of H_1 at significance level α" if the p-value is less than the desired significance level α (often $\alpha = 0.05$).
2. "Fail to reject H_0 in favor of H_1 at significance level α" otherwise.

Method. For tests about a proportion, the null hypothesis should be $H_0 : \theta = \theta_0$ and our test statistic is the sample total T. Under H_0, $T \sim \text{binom}(n, \theta_0)$.

6. A July, 2018 NPR/IPSOS poll asked respondents if they support or oppose "building a wall or fence along the entire U.S./Mexico border." Let θ be the proportion of the population in question (e.g. Midwesterners) that **oppose** building a wall or fence. Test $H_0 : \theta = 0.5$ against $H_1 : \theta > 0.5$ for the following populations. State your p-values and conclusions clearly.

a) 115 of 217 people in the Midwest oppose the wall

b) 150 of 264 people in the West oppose the wall

c) 198 of 401 people in the South oppose the wall

7. The article "Analysis of Reserve and Regular Bottlings: Why Pay for a Difference Only the Critics Claim to Notice?" reported on an experiment to determine if wine tasters could correctly distinguish between reserve and regular versions of a wine. In each trial tasters were given 4 indistinguishable containers of wine, two of which contained the regular version and two of which contained the reserve version of the wine. The taster then selected 3 of the containers, tasted them, and was asked to identify which one of the 3 was different from the other 2. In 855 trials, 346 resulted in correct distinctions. Does this provide compelling evidence that wine tasters can distinguish between regular and reserve wines?

a) Start with a null hypothesis that the tasters can't distinguish between the wines. State this as a hypothesis about the proportion of times the tasters correctly identify the odd wine out.

b) State an alternative hypothesis.

c) Calculate the observed significance level of the experimental results.

d) What is your conclusion? Any additional comments/thoughts?

Method. For tests about the mean of a population (with unknown variance), the null hypothesis should be $H_0 : \mu = \mu_0$ and our test statistics are given in column two of the table on page 1. Possible alternative hypotheses:

i) $H_1 : \mu \neq \mu_0$
ii) $H_1 : \mu > \mu_0$
iii) $H_1 : \mu < \mu_0$

8. The mean unrestrained compressive strength for a random sample of 45 specimens of a particular type of brick was 3107 psi, and the sample standard deviation was 188 psi. Does the data indicate that the true average unrestrained compressive strength is less than the design value of 3200? Assume the population is normally distributed and test $H_0 : \mu = 3200$ against $H_1 : \mu < 3200$ at a significance level of 0.05.

Definition. When conducting a hypothesis test there are two types of error:

(1) **Type I error** is rejecting H_0 when H_0 is actually true.
(2) **Type II error** is failing to reject H_0 when H_0 is actually false.

The **significance level** α of a test is the probability of a type I error.

Ideally, the probability of both types of error would be small. Unfortunately, decreasing the probability of type I error (by lowering the threshold for significance from 0.05 to 0.01, say) increases the probability of type II error (and vice versa). Since we control α, the usual strategy is to use the largest acceptable value for α (this means the probability of a type II error is as small as possible).

9. The EPA has determined that the Maximum Contaminant Level Goal (MCLG, "The level of a contaminant ... below which there is no known or expected risk to health") for nitrates in drinking water is 10mg/L. Imagine you have been hired by the City of Spokane to monitor drinking water safety. Your plan is to collect a random sample of water from 25 different sources and test the samples for nitrate levels, then use the sample mean to determine if Spokane's drinking water is safe.

a) State your null and alternative hypotheses in both non-mathematical language and as $H_0 : \mu = \_\_\_$ with a corresponding H_1.

b) State in plain language what a Type I error would be for your hypotheses (e.g. if you were explaining your results to the Mayor).

c) Repeat for a Type II error.

d) What significance level do you think makes sense for the test?

e) The EPA statement on the effect of nitrate contamination: "Infants below the age of six months who drink water containing nitrate in excess of the MCL could become seriously ill and, if untreated, may die. Symptoms include shortness of breath and blue-baby syndrome." Does this change the significance level you want to use for your test? Does this change how you want to set up the null and alternative hypotheses?