Hypothesis Tests & Confidence Intervals

Dave Zes

All Hypothesis Tests and Confidence Intervals do the same thing. A Hypothesis Test assumes the Claim (the Null Hypothesis, the status quo), and sees if the Observation is unusual in favor of the Alternative Hypothesis. A p-value is the probability of the observation occurring, or something more extreme in favor of the alternate hypothesis, assuming the Null is true. A sufficiently small p-value (say < 0.05) will advocate the alternative (and reject the Null). A Confidence Interval assumes the Observation is right, and sees if the Claim is unusual. If the Claim falls outside the interval, the Null is rejected.

Ask: Is the variable a proportion, or a quantitative variable? Are you asked for a confidence interval, or a test of hypothesis?

	Confidence Interval	Hypothesis Test
Quantitative variable w/ σ	$\overline{X} \pm z^{*}\sigma/\sqrt{n}$	$z = \left(\overline{X} - \mu_{_{H0}}\right)/(\sigma/\sqrt{n})$
Quantitative variable, no σ	$\overline{X} \pm t^{*}_{n-1}s/\sqrt{n}$	$t_{n-1} = \left(\overline{X} - \mu_{_{H0}}\right)/(s/\sqrt{n})$
Proportion	$\widehat{p} \pm z^{*}\sqrt{\widehat{p}\,\widehat{q}/n}$	$z = (\widehat{p} - p_{_{H0}})/\sqrt{p_{_{H0}}\,q_{_{H0}}/n}$
Difference in means	$\overline{X}_1 - \overline{X}_2 \pm t^{*}_{df}\mathsf{SE_i}\left[\overline{X}_1 - \overline{X}_2\right]$	$t^{*}_{df} = (\overline{X}_1 - \overline{X}_2)/\mathsf{SE_i}\left[\overline{X}_1 - \overline{X}_2\right]$
Difference in props	$\widehat{p}_1 - \widehat{p}_2 \pm z^{*}\mathsf{SE_p}\left[\widehat{p}_1 - \widehat{p}_2\right]$	$z = \left(\widehat{p}_1 - \widehat{p}_2 - (p_1 - p_2)\right)/\mathsf{SE_i}\left[\widehat{p}_1 - \widehat{p}_2\right]$

Table 1

n is the sample size.
n_1 is the sample size of sample number 1.
$s = \sqrt{\sum_i (X_i - \overline{X})^2/(n-1)}$ is the sample standard deviation.
σ is the population standard deviation.
$\overline{X}$ is the sample average (quantitative variable).
$\widehat{p}$ is the sample proportion.
μ_{H0} is the population average assuming the null hypothesis.
p_{H0} is the population proportion assuming the null hypothesis.
z^{*} is the z-score that corresponds to our desired confidence — virtually always 2-tailed.
t^{*} is the t-score that corresponds to our desired confidence — virtually always 2-tailed.
df is the degrees of freedom for a two-mean test. If equal variance assumed, approximately $df = n_1 + n_2 - 2$, otherwise apprx. $df = min(n_1, n_2)$
μ is the true, or hypothetically true, population average.
p is the true, or hypothetically true, population proportion.

$\mathsf{SD}\left[\,\overline{X}\,\right] = \sigma/\sqrt{n}$
$\mathsf{SE}\left[\,\overline{X}\,\right] = s/\sqrt{n}$
$\mathsf{SD}\left[\widehat{p}\,\right] = \sqrt{p_{H0}\, q_{H0}/n}$
$\mathsf{SE}\left[\widehat{p}\,\right] = \sqrt{\widehat{p}\,\widehat{q}/n}$

$$\mathsf{SE_p}\left[\widehat{p}_1 - \widehat{p}_2\right] = \sqrt{\frac{\widehat{p}_p(1-\widehat{p}_p)}{n_1} + \frac{\widehat{p}_p(1-\widehat{p}_p)}{n_2}}$$

where $\widehat{p}_p = (\widehat{p}_1 n_1 + \widehat{p}_2 n_2)/(n_1 + n_2)$

$$\mathsf{SE_i}\left[\widehat{p}_1 - \widehat{p}_2\right] = \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$$

$$\mathsf{SE_i}\left[\overline{X}_1 - \overline{X}_2\right] = \sqrt{\frac{s_1}{n_1} + \frac{s_2}{n_2}}$$

- ☛ All these methods are approximations.
- ☛ All these approximations require that the sample be *representative* of the population.
- ☛ When X is quantitative and σ is given, if X's distribution is close to normal, OR if n is sufficiently large (30 or more for slightly skewed data, perhaps 100 or more for more heavily skewed data), AND when our sample is no more than 10% of the population, then the approximation is very good.
- ☛ When X is quantitative and σ is not given (t-test), then X's distribution must be close to normal, or the approximation will be poor.
- ☛ When we deal with a proportion, p or $\widehat{p}$, when we pass the success failure rule ($np \geq 10$ and $nq \geq 10$, or $n\widehat{p} \geq 10$ and $n\widehat{q} \geq 10$), AND when our sample is no more than 10% of the population, the approximation is good.
- ☛ Always remember, when dealing with a proportion, when n is small, we can test a hypothesis by directly calculating the probability of what we've observed using the binomial distribution.

1 EXAMPLES

❶ A friend tells you he has a trick coin that comes up heads 80% of the time. You flip the coin 4 times and get 3 tails and 1 heads. What do we think of what we've been told? If we assume what we're told, then the probability of what we've observed, or something more contradictory to the claim, is $P = P(3 \text{ tails}) + P(4 \text{ tails}) = 4 \times (\frac{1}{5})^3 \times (\frac{4}{5})^1 \quad + \quad (\frac{1}{5})^4 \times (\frac{4}{5})^0 = 0.0272$. This is fairly strong evidence against the claim.

❷ A local newspaper reports that in their city, 20% of all teenagers play soccer in a league. In a representative sample of a 110 teenagers from this city, 11 say they're in a soccer league. What do we think of the newspaper's info? We'll perform a hypothesis test, that is, we'll assume the newspaper is correct, then see how improbable our observation is. We have $z = (\widehat{p} - p_{H0})/\sqrt{p_{H0} \cdot q_{H0}/n} = (0.10 - 0.20)/\sqrt{0.20 \cdot 0.80/110} = -2.62$. Such a z-score argues against the paper's claim. In a two-tailed test, the proportion we observed, or something more extreme, should only happen with probability 0.0088.

❸ A new ointment causes side effects 36 times on 200 representative patients. If the standard rate of side effects for this class of ointment is 22.5%, and we set significance at 5%, can we claim the new ointment is a significant improvement? We set $H_0 : p = 0.225$, and since we're only interested in betterness, that is, a lower rate, we set $H_A : p < 0.225$. $z = (\widehat{p} - p_{H0})/\sqrt{p_{H0} \cdot q_{H0}/n} = (0.18 - 0.225)/\sqrt{0.225 \cdot 0.775/200} = -1.52$. For a one-tailed test at $\alpha = 0.05$, the critical cut-off is $z* = -1.64$. Our observation does not sufficiently contradict the null hypothesis, so there's nothing special about the side effect rate of the new ointment.

❹ The national average for a standardized test is 68.51, with a standard deviation of 7.87. At the Flintstone Acadamy, 128 representative students took the test and scored an average of 63.10, with a standard deviation of 7.32. If we want to show that Flintstone Acadamy is superior to the national average, what's our p-value? Since we are given the population standard deviation, σ we perform a z-test. $z = (63.10 - 68.51)/(7.87/\sqrt{128}) = -7.78$. Since we are interested in the prospect that μ at the Flintstone Acadamy is greater than μ nationally, we want the area of the right tail of the Z-distribution: $P(Z > -7.78) \approx 1$. There is essentially no evidence that the Flintstone Academy is better.

❺ Suppose Jim and Jane are the only two candidates in a political race. If we set significance at 5%, how many representative voters would we need to poll to be able to detect that Jane's true support is 51% or better, as opposed to 50%? We can set this up as a one-tailed CI, or test of hypothesis. We'll do the latter. For a one-tailed test at $\alpha = 0.05$, our critical cut-off, $z* = 1.64$. So we have that $1.64 = 0.01/\sqrt{p_{H0} \cdot q_{H0}/n} = 0.01/\sqrt{0.50 \cdot 0.50/n} \iff 164 = 1/\sqrt{0.50 \cdot 0.50/n} \iff \sqrt{0.50 \cdot 0.50/n} = 1/164 \iff 1/(4n) = 1/26896 \iff n = 6724$. Note that the equation for this can be written as $n = (z^*)^2 \cdot p \cdot q/\delta^2$.

Same scenario. How many representative voters would we need to poll to be able to detect that Jane's true support is 53% or better? $n = (1.64)^2 \cdot 0.5 \cdot 0.5/(0.03)^2 \iff n = 747$.

❻ A consumer rights group tests a representative sample of 18 connoisseur chocolate bars and finds that their average amount of cocoa is 16.28 grams, with a standard deviation of 115 milligrams. The manufacturer claims that these bars are made with 16.50 grams of cocoa. What do we conclude? We do not know the population standard deviation of the amount of cocoa in each bar; in lieu we have only the sample standard deviation. So, we have $t = (16.28 - 16.50)/(0.115/\sqrt{18}) = -8.1$. We are naturally only interested in customers receiving less than what they expect, so our H_A : $\mu < 16.50$ grams. So, $P(T < -8.1) = 0$. This p-value is, of course, less than any reasonable cut-off, hence we conclude that that manufacturer is misrepresenting their product.

❼ The diameter of a Gizmo X main lever pin is a normally distributed random variable with mean 10.0000 mm and standard deviation of 0.0012 mm. What is the probability that in a sample of 5, all 5 have a diameter less than 10.0018? By close inspection, we can see that we're interested in the probabilty of a part's diameter being less than one-and-a-half standard deviations above the mean: $Z = (X - \mu)/\sigma = (10.0018 = 10.0000)/0.0012 = 1.5$. And so, $P(Z < 1.5) = 0.9332$. The probability that 5 of 5 have diameters less than 10.0018 is $(0.9332)^5 = 0.7077$

What's the probability that in a sample of 200, 92.00% or more will have a diameter of less than 10.0018? We have $z = (\widehat{p} - p_{H0})/\sqrt{p_{H0} \cdot q_{H0}/n} = (92.00 - 0.9332)/\sqrt{0.9332 \cdot 0.0668/200} = -0.748$. So, $P(Z > -0.748) = 0.773$.

What's the probability that the average diameter will be more than 10.0018? We use a z-test: $z = (10.0018 - 10.0000)/(0.0012/\sqrt{200}) = 21.2$. So $P(Z > 21.2) = 0$.