M358K: November 19th, 2021.

t-procedures [cont'd].

Review: We have a normal population w/ both the mean parameter μ and the standard deviation σ unknown.

The procedures also work for large samples even from skewed distributions.

Let $X_1, X_2, \dots, X_n$ be a random sample.

Define:

- the sample mean: $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$
- the sample variance: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
- the sample std deviation: S

Define our t-statistic:

$$\boxed{T := \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(df = n-1)}$$ ☆

To construct a confidence interval @ the confidence level C, we use

$$\bar{X} - t^* \cdot \frac{S}{\sqrt{n}} < \mu < \bar{X} + t^* \cdot \frac{S}{\sqrt{n}}$$

w/ probability C.

The confidence interval (once the data are gathered and analyzed) is:

$$\boxed{\mu = \bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}}}$$

Hypothesis test for μ.

Hypotheses:

$$H_0: \mu = \mu_0 \quad \text{vs.} \quad H_a: \begin{cases} \mu < \mu_0 \\ \mu \neq \mu_0 \\ \mu > \mu_0 \end{cases}$$

Test statistic (under the null hypothesis).

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t(df = n-1)$$

Let t be the observed value of the test statistic, i.e.,

$$\boxed{t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}} \quad \star$$

For the p-value, we calculate the probability of observing what we observed or something more extreme.

IF $H_a: \mu < \mu_0$, then p-value $= \mathbb{P}[T < t]$

IF $H_a: \mu \neq \mu_0$, then p-value $= \mathbb{P}[T < -|t|] + \mathbb{P}[T > |t|]$

$$= 2 \cdot \mathbb{P}[T < -|t|] = 2 \cdot \mathbb{P}[T > |t|]$$

IF $H_a: \mu > \mu_0$, then p-value $= \mathbb{P}[T > t]$

for $\boxed{T \sim t(df = n-1)}$

If a significance level α is given, then we can construct a rejection region (RR).

$$RR = \begin{cases} (-\infty, -t_{\alpha,n-1}] \\ (-\infty, -t_{\frac{\alpha}{2},n-1}] \cup [t_{\frac{\alpha}{2},n-1}, +\infty) \\ [t_{\alpha,n-1}, +\infty) \end{cases}$$

with $\boxed{t_{\alpha,n-1}}$ being a value s.t. $\mathbb{P}[T > t_{\alpha,n-1}] = \alpha$ w/ $T \sim t(df=n-1)$

or, in R, `qt(1-α, df=n-1)`.

Decision: If the observed value of the test statistic t falls in the RR, then we reject the null. Otherwise, we fail to reject null.

Example. [Ramachandran·Tsokos]

A manufacturer of fuses claims that w/ a 20% overload, their fuses blow in less than 10 minutes on average.

To test this claim, a random sample of 20 fuses is gathered an subjected to the 20% overload. The times it took them to blow had the sample mean of $\boxed{10.4}$ minutes and the sample std deviation of $\boxed{1.6}$ minutes.

Assume that the times come from a normal dist'n.

Do the data support or refute the manufacturer's claim?

→: X... the "reaction" time, i.e., the population dist'n.

$X \sim Normal(mean = \mu, sd = \sigma)$

$\boxed{H_0: \mu = 10}$ vs. $H_a: \mu > 10$

$n = 20, \quad \bar{x} = 10.4, \quad s = 1.6$

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{10.4 - 10}{\frac{1.6}{\sqrt{20}}} = \boxed{1.118034}$$

p-value: 1-pt(t, df=19) = 0.1387451

=> fail to reject (at any "reasonable" significance level).

Using the t-table, get that the p-value is between 10% and 15%