t tests

Psychology 3256

Introduction

- Remember, to find the probability of a given value of a variable, if the variable is normally distributed, you just turn it into a z score
- $z = (x-\mu/\sigma)$
- Look it up in a table, and we are in business

As interesting as this is...

- We are usually dealing with means, not with individual values
- Instead of knowing about the distribution of x, we usually care about the distribution of $\bar{x}$

The Central Limit Theorem

- Given a population with a mean μ and a variance σ^2 the sampling distribution of the mean will have a mean of μ ($\mu_{\bar{x}} = \mu$) and a variance of σ^2/n. As n increases this distribution approaches normal **no matter what the shape of the parent population distribution**

Oooh the power

- The population distribution shape does not matter.
- What matters is random sampling (though not as much as you might think)
- so, to find out $p(\bar{x})$ we just use our old friend the z score

So it changes a bit

$$z = \frac{x - \mu}{\sigma} \rightarrow z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

an example

- Let’s say 25 subjects are given an ‘IQ Improvement course’ and have their IQs tested after the course. They end up with a mean of 110.
- IQ in the population has a μ of 100 and a σ of 15

Do the plusses and take aways...

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$z = \frac{110 - 100}{\frac{15}{\sqrt{25}}}$$

$$z = \frac{10}{\frac{15}{5}}$$

$$z = \frac{10}{3} = 3.33 \rightarrow p(z > 3.33) < .00043$$

There is a problem here

- We don't typically know σ for a population
- Well $E(s^2)=\sigma^2$ right?
- We could just sub s^2 for σ^2 but now we have two sampling distributions
- The sampling distribution of s^2 will change depending on n
- so we cannot use z

So we use t

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

remember

- t changes depending on the number of observations
- degrees of freedom to estimate σ^2 (by calculating s^2) to be more precise
- Pretty powerful technique, you don't have to know anything about the population!

Pairs of observations

- Can also be used for pairs of observations
- Same with matching
- Be careful, now subjects are different on everything if you matched

$$t = \frac{\bar{x}_d}{\frac{s_d}{\sqrt{n}}}$$

What if your pairs are not matched?

- Aha, two populations (maybe)
- well then they hypotheses are as follows
- H_0 $\mu_1 = \mu_2$
- Ha $\mu_1 \neq \mu_2$

The original t formula

Statistic ↓   **H_0** ↓

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

← **Error**

Figure it out

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{error}$$

practically....

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{error}$$

Now for the error

- variances have to be weighted

$$\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

So our formula is

- So really we have the same formula, we just subbed different values in for the statistic, the Ho and the error

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

If the variances are not equal things change

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

How many df?

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

Assumptions

- SRS
- Independence of observations
- homogeneity of variance