UNIVERSITY OF TEXAS AT AUSTIN

Problem Set # 20

The t-procedure: two means

20.1. **Two-sample t.**

Problem 20.1. An instructor is teaching two sections of the same basic statistics course. The instructor is giving the same exams, homework assignments, and quizzes in both sections. Which $t-$procedure should be used to determine if there is a difference in the academic performance between the two course sections?

(a) One-sample $t-$test.
(b) Matched-pairs $t-$procedure.
(c) Two-sample $t-$test.
(d) None of the above.

Problem 20.2. This is an excerpt from findings of an educational study:

A study was done to determine whether there is a difference in the amount of time (in hours) that graduate students versus undergraduate students spend on the Internet per day. Five undergrads and five grad students were polled.

(i) Is the alternative hypothesis one-sided or two-sided?
(ii) A $t-$score for the data gathered was calculated to be 1.6664. Would you say that there is a significant difference in the amount of time that graduate and undergraduate students spend on the Internet? No!

p-value = 2 * pt(-1.6664, df=4) = 0.1709623

Problem 20.3. (5 points)

There is a dispute about salaries of male versus female elves. The North Polar Bear collected the following data:

- the total number of male elves is 121;
- the total number of female elves is 100;
- the average salary of a male elf is $10,000$ candy canes;
- the average salary of a female elf is $12,000$ candy canes;
- the sample standard deviation of the salaries of male elves is 50;
- the sample standard deviation of the salaries of female elves is 132.

Assume independence between the salaries of individual elves.

Let μ_m denote the population mean for the distribution of the male elves' salaries and let μ_f denote the population mean for the distribution of the female elves' salaries. We wish to test:

$$H_0 : \mu_m = \mu_f \quad vs. \quad H_a : \mu_m \neq \mu_f.$$

What is the $p-$value associated with our data?

a.: About 0.
b.: About 0.01.
c.: About 0.025
d.: About 0.04.
e.: None of the above.

$\rightarrow$: Two-sided t-test

$$t = \frac{\bar{x}_m - \bar{x}_f}{\boxed{\sqrt{\frac{s_m^2}{n_m} + \frac{s_f^2}{n_f}}}} = \frac{10000 - 12000}{\sqrt{\frac{50^2}{121} + \frac{132^2}{100}}} \Rightarrow$$

$$\boxed{t = -143.26}$$ $\Rightarrow$ the p-value is negligible

$$df = \min(n_1, n_2) - 1 = 99$$

Problem 20.4. Let the population distributions be normal with unknown parameters. Assume that sample data, based on two independent samples of size 25, give us $\boxed{\bar{x}_1 = 505, \bar{x}_2 = 515}$, $s_1 = 23$, and $s_2 = 28$.

(i) What is a 95%-confidence interval (use the conservative value for the degrees of freedom) for the difference between the two population means?

(ii) Based on the confidence interval, we can conclude at the 5% significance level that there is no difference between the two population means. *True or false?*

(iii) The margin of error for the difference between the two sample means would be smaller if we were to take larger samples. *True or false?*

(iv) If a 99% confidence interval were calculated instead of the 95% interval, it would include more values for the difference between the two population means. *True or false?*

20.2. **Pooled t (optional content from Subsection 7.3.4).**

Problem 20.5. The **pooled** two-sample $t-$procedure can be used when ...

(a) you can assume the two populations have equal variances
(b) you can assume the two populations have equal means
(c) the sample sizes are equal
(a) None of the above

Problem 20.6. Let n_1 and n_2 denote the sample sizes of each group. The pooled two-sample $t-$procedure is based how many degrees of freedom?

(a) $n_1 + n_2 + 2$
(b) $n_1 + n_2 - 2$
(c) $n_1 + n_2 - 1$
(d) $n_1 + n_2$
(e) None of the above.

Problem 20.7. A study was done to determine if students learn better in an online basic statistics class versus a traditional face-to-face (f2f) course. A random sample of 12 students in an online course and 15 students in an f2f course was taken.

(i) Let μ_{new} denote the population mean score for the online statistics class and let μ_{old} denote the population mean score for the face-to-face statistics class. What are the hypotheses being tested?

(ii) We decide it is appropriate to use the pooled $t-$procedure. What is the number of degrees of freedom you are going to use?

---

i. $\boxed{\text{pt.estimate} \pm \text{margin.of.error}}$

- pt.estimate for $\mu_1 - \mu_2$: $\bar{x}_1 - \bar{x}_2 = 505 - 515 = -10$
- margin of error = $t^* \cdot$ std error
  - std error $= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{23^2}{25} + \frac{28^2}{25}} =$

    $= \frac{\sqrt{23^2 + 28^2}}{5} = 7.247$
  - $t^* = ?$ df = 25 − 1 = 24, upper tail probab. = 0.025 $\Rightarrow$ $t^* = 2.064$

    t tables

$$\Rightarrow \mu_1 - \mu_2 = -10 \pm 2.064(7.247) = -10 \pm 14.96$$