Lesson 8: Student's t

Typically the variance σ^2 is unknown.

Use s^2 to estimate σ^2 from the data.

$$s^2 = \frac{1}{n-1}\sum(y_i - \bar{y})^2$$

$(n-1)s^2/\sigma^2$ has a chi squared distribution with $n-1$ degrees of freedom.

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

Lesson 9: Comparing Two Means

Observations of Motor Types

Type A	Type B
64.9	62.7
62.1	61.8
60.7	63.3
63.8	65.2
69.7	60.8
62.9	

Type A	Type B
$\bar{y}_A = 64.02$	$\bar{y}_B = 62.76$
$n_A = 6$	$n_B = 5$
$s_A^2 = 9.81$	$s_B^2 = 2.75$
$\nu_A = 5$	$\nu_B = 4$

The null hypothesis is that there is no difference between A and B.

$$H_0 : \mu_A - \mu_B = 0$$

Need to pool the variance estimates.

$$s^2 = \frac{\nu_A s_A^2 + \nu_B s_B^2}{\nu_A + \nu_B} = 6.67$$

Calculate the t statistic.

$$t_9 = \frac{(\bar{y}_A - \bar{y}_B) - (\mu_A - \mu_B)}{\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right) s^2}} = \frac{1.26 - 0}{\sqrt{\left(\frac{1}{6} + \frac{1}{5}\right) 6.67}} = 0.8$$

Is this value of t a rare event? No, it is not. Hence the data do not contradict the hypothesis that $\mu_A = \mu_B$.

Now compute an interval estimate for $\mu_A - \mu_B$. The formula for the interval estimate is

$$\bar{y}_A - \bar{y}_B \pm t_{\nu,\alpha/2} \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right) s^2}$$

Plug in the numbers.

$$1.26 \pm 2.26 \sqrt{\left(\frac{1}{6} + \frac{1}{5}\right) 6.67}$$

The result is

$$1.26 \pm 3.56$$

Hence

$$-2.30 \leq (\mu_A - \mu_B) \leq 4.82$$

"All values of $\mu_A - \mu_B$ within these limits are not contradicted by the data."

$$P[-2.30 \leq (\mu_A - \mu_B) \leq 4.82] = 0.95$$

Some statements about the data.

$$E(\bar{y}_A - \bar{y}_B) = \mu_A - \mu_B$$
$$V(\bar{y}_A - \bar{y}_B) = \left(\frac{1}{n_A} + \frac{1}{n_B}\right) \sigma^2$$

It turns out that $n_A = n_B$ minimizes the variance.

Three important assumptions about the data for a t statistic.

- Independence
- Normality

- Homogeneous Variance

Homogenous variance is also known as homoscedasticity.

Independence is ensured by random selection of motors.

The t statistic is a "robust" statistic. The t statistic still works well when the requirements of normality and homoscedasticity are relaxed.

Lesson 10: The ANOVA Case

8
12
6
10
5
3
1
7
3
5

$$\sum y = 60$$
$$\bar{y} = 6$$
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n-1}\left[\sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2\right]$$
$$\sum y^2 = 462$$
$$s^2 = 102/9$$

Model

$$y_i = \mu + \epsilon_i$$

with

$$\epsilon_i \sim N(0, \sigma^2)$$

ANOVA table:

											SSq	df
y_i	8	12	6	10	5	3	1	7	3	5	462	10
$\bar{y}$	6	6	6	6	6	6	6	6	6	6	360	1
$y_i - \bar{y}$	2	6	0	4	-1	-3	-5	1	-3	-1	102	9

Observe that SSq is partitioned, i.e., $462 = 360 + 102$.

Observe that $s^2 = 102/9 = \text{SSq/df}$.

New information about the data.

A	B
8	5
12	3
6	1
10	7
	3
	5

$$\bar{y}_A = 9 \quad \bar{y}_B = 4$$

$$s_A^2 = 20/3 \quad s_B^2 = 22/5$$

Pooled estimate of the variance:

$$s^2 = 42/8, \quad \nu = 8$$

Hypothesize that the means of A and B are the same.

$$H_0 : \mu_A - \mu_B = 0$$

Compute the t statistic.

$$t = \frac{(\bar{y}_A - \bar{y}_B) - (\mu_A - \mu_B)}{\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right) s^2}} = 3.38$$

Is this an unusual value of t when H_0 is true?

The critical value of the t_8 distribution is 2.308.

Since t is greater than 2.308 we reject the null hypothesis that $\mu_A = \mu_B$.

New model:

$$y_{ij} = \mu + \tau_j + \epsilon_{ij} \qquad \epsilon_{ij} \sim N(0, \sigma^2)$$

New ANOVA table:

											SS	df
y_{ij}	8	12	6	10	5	3	1	7	3	5	462	10
$\bar{y}$	6	6	6	6	6	6	6	6	6	6	360	1
τ_j	3	3	3	3	-2	-2	-2	-2	-2	-2	60	1
Residuals	-1	3	-3	1	1	-1	-3	3	-1	1	42	8

Variance estimate can be computed from ANOVA table:

$$s^2 = SS/df = 42/8$$

More compact ANOVA table:

Source	SS	df
$\sum y^2$	462	10
CF	360	1
TSS	60	1
RSS	42	8

I	II
74.0	73.2
68.8	68.2
71.2	70.9
74.2	74.3
71.8	70.7
66.4	66.6
69.8	60.5
71.3	70.8
69.3	68.8
73.6	73.3