Week 10

Monday, 8 January 2024 9:07 AM

## TWO MEANS → Quantitative data

**Two means** $$(\bar{y}_1 - \bar{y}_2) \pm t^*_{n_1+n_2-2} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

| Group 1 | Group 2 | Group 3 |
| --- | --- | --- |
| Sample size, $n_1$ | $n_2$ | $n_3$ |
| Average, $\bar{y}_1$ | $\bar{y}_2$ | $\bar{y}_3$ |
| Standard deviation, $s_1$ | $s_2$ | $s_3$ |

Compare the results

Compare the results

Compare the results

$\boxed{\text{ANOVA}}$ → Analysis of Variance → 3 or More groups

F-model or F-test by Sir Ronald Fisher

The hypotheses are of the form:

$H_0$: $\mu_1 = \mu_2 = \ldots = \mu_k$ (all means are equal)

$H_A$: not all means are equal (or, at least two means differ)

Example : Examples include comparing the hours of pain relief for four different pain relief drugs, or comparing exam marks across three different campuses.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_A$: not all means are equal

→ $H_0: \mu_1 = \mu_2 = \mu_3$

$H_A$: not all means are equal

And it's the central idea of the *F*-test.

- We compare the differences *between* the means of the groups with the variation *within* the groups.

Error mean Square, $MS_E$ → Variation within the groups

Treatment mean Square, $MS_T$ → Variation between the groups

$$F\text{-statistic} = \frac{MS_T}{MS_E}$$

F-statistic is Close to 1, P-value is HIGH → FAIL to Reject Null Hypotheses

F-statistic is more than 1, P-value is LOW → Reject $H_0$

For $MS_T$ → $df = k - 1$

$k$ → number of groups

For $MS_E$ → $df = k(n-1)$

$n$ → sample sise

**Question 2: ANOVA conclusion**

The null hypothesis associated with the soap washing example used in table above is that washing with all soaps results in the same average number of bacteria colonies. Or, $H_0$: $\mu_1 = \mu_2 = \mu_3 = \mu_4$.

Analysis of Variance Table

| Source | Sum of Squares | DF | Mean Square | F-ratio | P-value |
|---|---|---|---|---|---|
| Soaps | 29882 | 3 | 9960.64 | 7.0636 | 0.0011 |
| Error | 39484 | 28 | 1410.14 | | |
| Total | 69366 | 31 | | | |

From the ANOVA table above and using a significance level of 5%, give the test statistic, the P-value and your conclusion for the test.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_A$: All means are not equal

F-statistic = 7.0636 P-value = 0.0011

P-value is Low, Reject $H_0$
$\therefore$ All means are not equal

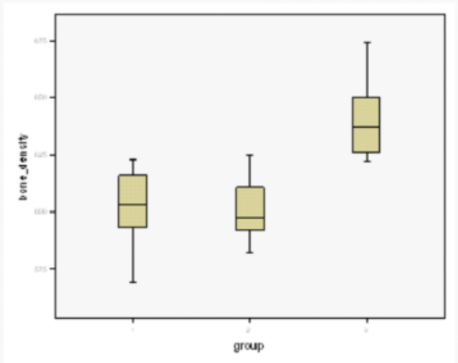
## Bonferroni Multiple Comparisons

If we reject the overall Null Hypotheses, we do multiple comparisons

**Question 3 : Bonferroni's multiple comparisons example**

Quantitative variable

In a study to investigate the link between exercise and healthy bones, the bone density of growing rats was examined using jumping. There were three treatments: a control (no jumping), a low-jump group (jump height 30cm) and a high-jump group (height 60cm). The rats in the jump groups jumped 10 jumps per day, 5 days per week for 8 weeks. Data summaries and SPSS output of the ANOVA analysis with Bonferroni's multiple comparisons are given below.




| Group | mean | variance |
|---|---|---|
| 1: control | 602.2 | 322.2 |
| 2: low | 601.8 | 220.6 |
| 3: high | 638.7 | 275.3 |

**ANOVA**

Bone_density

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 8825.296 | 2 | 4412.648 | 16.294 | .000 |
| Within Groups | 7041.256 | 26 | 270.818 | | |
| Total | 15866.552 | 28 | | | |

**Multiple Comparisons**

Dependent Variable: bone_density

Bonferroni

| (I) group | (J) group | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | Upper Bound |
|---|---|---|---|---|---|---|
| 1 | 2 | .422 | 7.561 | 1.000 | -18.93 | 19.77 |
| | 3 | -36.478* | 7.561 | .000 | -55.83 | -17.13 |
| 2 | 1 | -.422 | 7.561 | 1.000 | -19.77 | 18.93 |
| | 3 | -36.900* | 7.360 | .000 | -55.73 | -18.07 |
| 3 | 1 | 36.478* | 7.561 | .000 | 17.13 | 55.83 |
| | 2 | 36.900* | 7.360 | .000 | 18.07 | 55.73 |

*. The mean difference is significant at the 0.05 level.

a) State the null and alternative hypotheses for the ANOVA test.
b) Give the test statistic and P-value.
c) Write a brief conclusion of the ANOVA analysis. Use α=0.05.
d) Summarise the results of Bonferroni's multiple comparisons.

(a) $H_0: \mu_1 = \mu_2 = \mu_3$

$H_A$: All means are not equal

(b) F-statistic = 16.294

p-value = 0.000

(c) p-value < 0.05

p-value is Low, Reject Null Hypotheses

(d) 601.8 602.2 638.7

Group 2 Group 1 Group 3

There is evidence that the mean bone density for the high jumping group is significantly higher than the low jumping and control groups. OR,

Mean bone density for group 1 and 2 is same and lower than group 3.

OR

Mean bone density for group 3 is higher than group 1 and 2

Problem solving task 3

CHI-SQUARE TEST

**Q3.1** Smarties (sugar coated chocolate confectionary) come in 8 colours – green, yellow, red, orange, pink, purple, blue and brown. You buy a bag containing 120 smarties to investigate the distribution of colours, and count 12 green, 14 yellow, 17 red, 15 orange, 16 pink, 17 purple, 11 blue and 18 brown smarties.

a) If smarties are packaged in equal proportions, how many of each colour would you expect in the bag?
b) To see if these results are unusual, should you perform a goodness-of-fit test or a test of independence?
c) State your hypotheses.
d) How many degrees of freedom are there? df = n − 1
e) Find $\chi^2$ and the P-value. Chi-square table

$H_0$: obsv. count = Exp. counts

$H_A$: obsv. counts ≠ Exp. counts

| | Obsv. | Exp. |
|---|---|---|
| Green | 12 | 15 |
| Yellow | 14 | 15 |
| Red | 17 | 15 |
| Orange | 15 | 15 |
| Pink | 16 | 15 |

c) State your hypotheses.
d) How many degrees of freedom are there? df = n − 1
e) Find $\chi^2$ and the P-value. Chi-square table
f) State your conclusion (use α = 0.05) in the context of the question.

H₀: obsv. counts = Exp. counts
$H_A$: obsv. counts ≠ Exp. counts

| | | |
|---|---|---|
| Orange | 15 | 15 |
| Pink | 16 | 15 |
| purple | 17 | 15 |
| blue | 11 | 15 |
| brown | 18 | 15 |

[1+1+1+1+3+1 = 8 marks]

$$\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp}$$

CHI-SQUARE TEST OF INDEPENDENCE

**Q3.2** The following table shows data on randomly selected data on security level of staff and position held. Are security level and the position held in the company independent?

| | Very Secure (obsv 1) | Secure (obsv 2) | Insecure (obsv 3) | |
|---|---|---|---|---|
| Non manager | 22 | 55 | 15 | 92 |
| Manager | 45 | 35 | 15 | 95 |
| | 67 | 90 | 30 | $\boxed{187}$ |

Exp 1 Exp 2 Exp 3

$$\text{Expected Count} = \frac{\text{row total} \times \text{Column total}}{\text{Total number of observations}}$$

a) Write appropriate hypotheses.
b) How many degrees of freedom are there? df = (R − 1)(C − 1)
c) Find $\chi^2$ and the P-value.
d) State your conclusion (use α = 0.05).

H₀: independent
$H_A$: dependent

$$\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp}$$

ONE MEAN

**Q3.3** During an angiogram, heart problems can be examined via a small tube (a catheter) threaded into the heart from a vein in the patient's leg. It is important that the company that manufactures the catheters maintains a diameter of 2.00 mm. A random sample of 36 catheters is taken and the average diameter is 2.03 mm with standard deviation 0.05 mm.

$\mu_0 = 2$
$n = 36$
$\bar{y} = 2.03$
$s = 0.05$

a) Create a 95% confidence interval for the mean diameter of catheters produced by the company.
b) Explain in context what your interval means.
c) Perform a hypothesis test to find out if the mean diameter of the catheters is significantly different to the required 2.00 mm. Use a significance level of 5% and give your conclusion.

$df = n - 1 = 36 - 1 = 35$

$t_{35}(95\%) = 2.03$

$H_0: \mu = 2$
$H_A: \mu \neq 2$

[4+2+ 6= 12 marks]

Confidence interval

Test statistic

One mean $\quad \bar{y} \pm t^*_{n-1} \times \frac{s}{\sqrt{n}} \qquad t = \frac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}}$

**Q3.4** Many drivers believe that they can get better gas mileage by using premium rather than regular gas. To test this we use 10 cars from a company fleet in which all the cars run on regular gas. Each car is filled first with either regular or premium gasoline decided by a coin toss, and the mileage for that tankful is recorded. Then the mileage is recorded again for the same cars for a tankful of the other kind of gasoline. We do not let the drivers know about this experiment. The results are provided in the table below:

PAIRED MEANS

**Miles per gallon**

| Car# | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Regular | 16 | 20 | 21 | 22 | 23 | 22 | 24 | 28 | 27 | 28 |
| Premium | 18 | 21 | 23 | 24 | 25 | 23 | 26 | 26 | 28 | 29 |
| d | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 |

$n = 10$

$\bar{d} = \frac{16}{10} = 1.6$

$s_d = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}}$

$df = n-1 = 10-1 = 9$

a) Do the use of premium gasoline differ significantly in the mileage? Carry out a hypothesis test (preferably with technology such as SPSS) using α=0.05 and write your conclusion.

b) Calculate a 90% confidence interval for the difference in mileage for the two different gasoline and interpret the interval.

$t_9(90\%) = 1.833$

**[5+5 = 10 marks]**

Confidence interval

<u>Paired means</u> $\bar{d} \pm t^*_{n-1} \times \frac{s_d}{\sqrt{n}}$

Test statistic

$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$

**Q3.5** A researcher wanted to see whether there is a significant difference in the blood pressure of men and women. He collects data from 23 men and 35 women. The average systolic blood pressure reading for women was 105 with a standard deviation of 12 whereas for the men the average was 115 with a standard deviation of 15. The data are summarised below:

TWO MEANS

| | | Mean | Std.Dev. |
|---|---|---|---|
| Group 1 | Men | 115 $\bar{y}_1$ | 15 $s_1$ |
| Group 2 | Women | 105 $\bar{y}_2$ | 12 $s_2$ |

sample size
$n_1 = 23$
$n_2 = 35$

$H_0: \mu_1 = \mu_2$
$H_A: \mu_1 > \mu_2$

a) Are systolic blood pressure readings for the men significantly higher than those of women? Carry out a hypothesis test. Use a significance level of 5% and use df = $n_1+n_2-2$. $= 23+35-2 = 56$

b) Create a 90% confidence interval for the difference in mean systolic blood pressure readings and interpret the interval.

c) Does the confidence interval confirm your answer to a)?
Explain.

**[6+5+1 = 12 marks]**

Confidence interval

<u>Two means</u> $(\bar{y}_1 - \bar{y}_2) \pm t^*_{n_1+n_2-2} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Test statistic

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

$t^*_{56}(90\%) = 1.671$