**Tutorial Stats Inference 2 - Solutions**

## 1. Primary School

A random sample of 25 ten year olds were taken from primary schools in each of Scotland, England, Northern Ireland and Wales, and their height was then measured in inches. Choose the appropriate method of analysis to determine if there were significant differences between the heights of 10 year olds in each country, and if so which countries were significantly different from which?

Let $\mu_i$ be the mean height of 10 year olds from country $i$.

Let s= Scotland, e=England, n=Northern Ireland and w=Wales.

Then

$$H_0: \mu_s = \mu_e = \mu_n = \mu_w$$

$$H_1: at\ least\ one\ \mu_i\ not\ equal$$

First we need to check the assumptions which are:

- The subjects have been selected at random from the m groups.
- The dependent variable/response is normally distributed in each group.
- The dependent variable has the same variance in each group (denoted as $\sigma^2$).

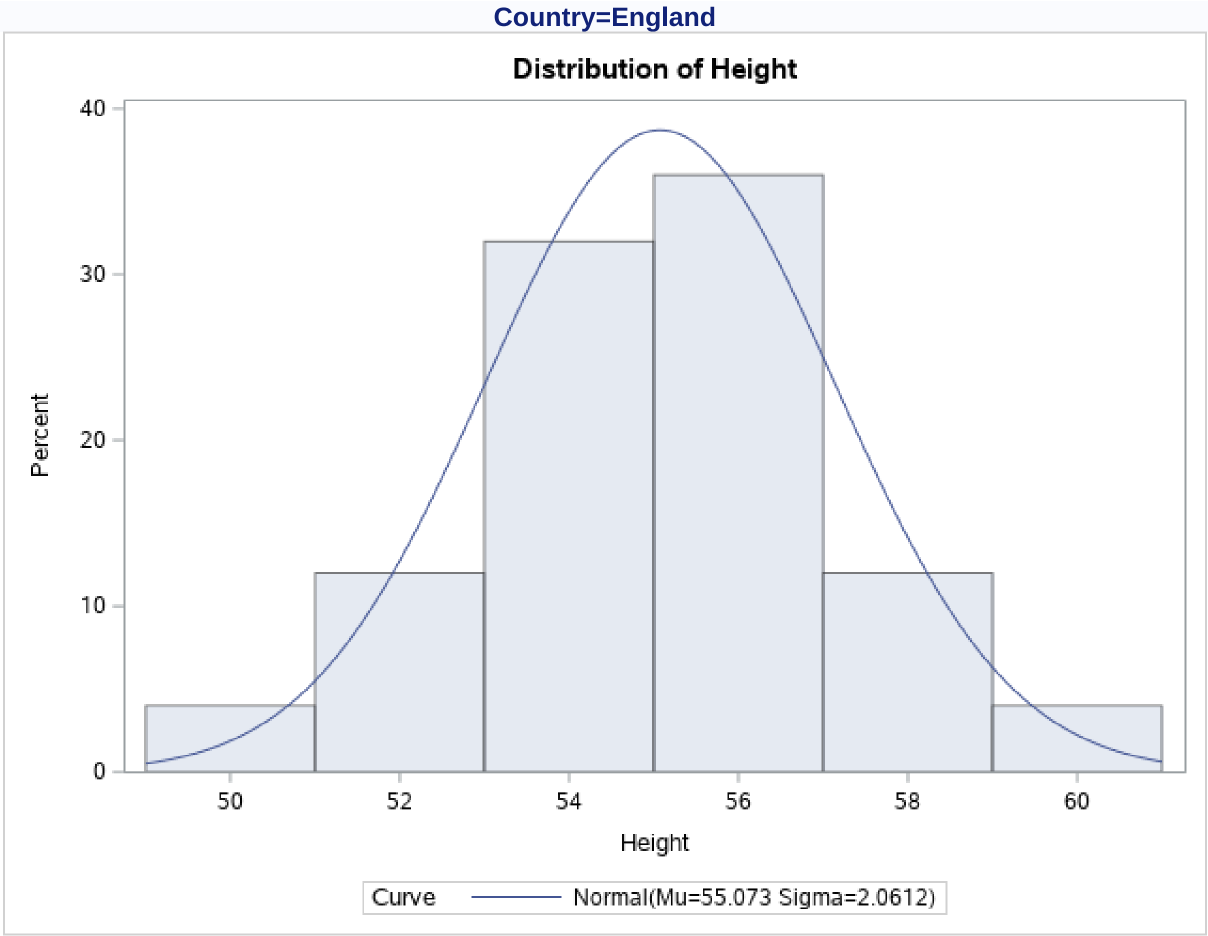
We are told that the subjects have been randomly selected therefore this assumption has been met. We therefore need to now check that height is normally distributed in each school, the null and alternative hypotheses being that:

$$H_0: \text{samples } x_s,\ x_e, x_n, x_w \text{came from a normally distributed population}$$

$$H_1: at\ least\ one\ sample\ did\ not\ come\ from\ a\ normally\ distributed\ population$$

This is the output table obtained for the school in England along with the corresponding histogram:

| **Goodness-of-Fit Tests for Normal Distribution** | | | | |
|---|---|---|---|---|
| **Test** | **Statistic** | | **p Value** | |
| Kolmogorov-Smirnov | D | 0.10685969 | Pr > D | >0.150 |
| Cramer-von Mises | W-Sq | 0.03659763 | Pr > W-Sq | >0.250 |
| Anderson-Darling | A-Sq | 0.22372461 | Pr > A-Sq | >0.250 |




Looking at the p-values for all tests for normality, no significant difference is found. Therefore we do not reject the null hypothesis, so there is no evidence to suggest that the distribution is not a normally distributed.

Carrying out this procedure for all other countries yields the same result, therefore we can conclude that we meet the assumption for normality, hence do not reject the null hypothesis.

The final assumption is that of equal variances, with the null and alternative hypothesis being:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ \dots\dots = \sigma_m^2$$

$$H_1: \sigma_i^2 \text{ are not all equal.}$$

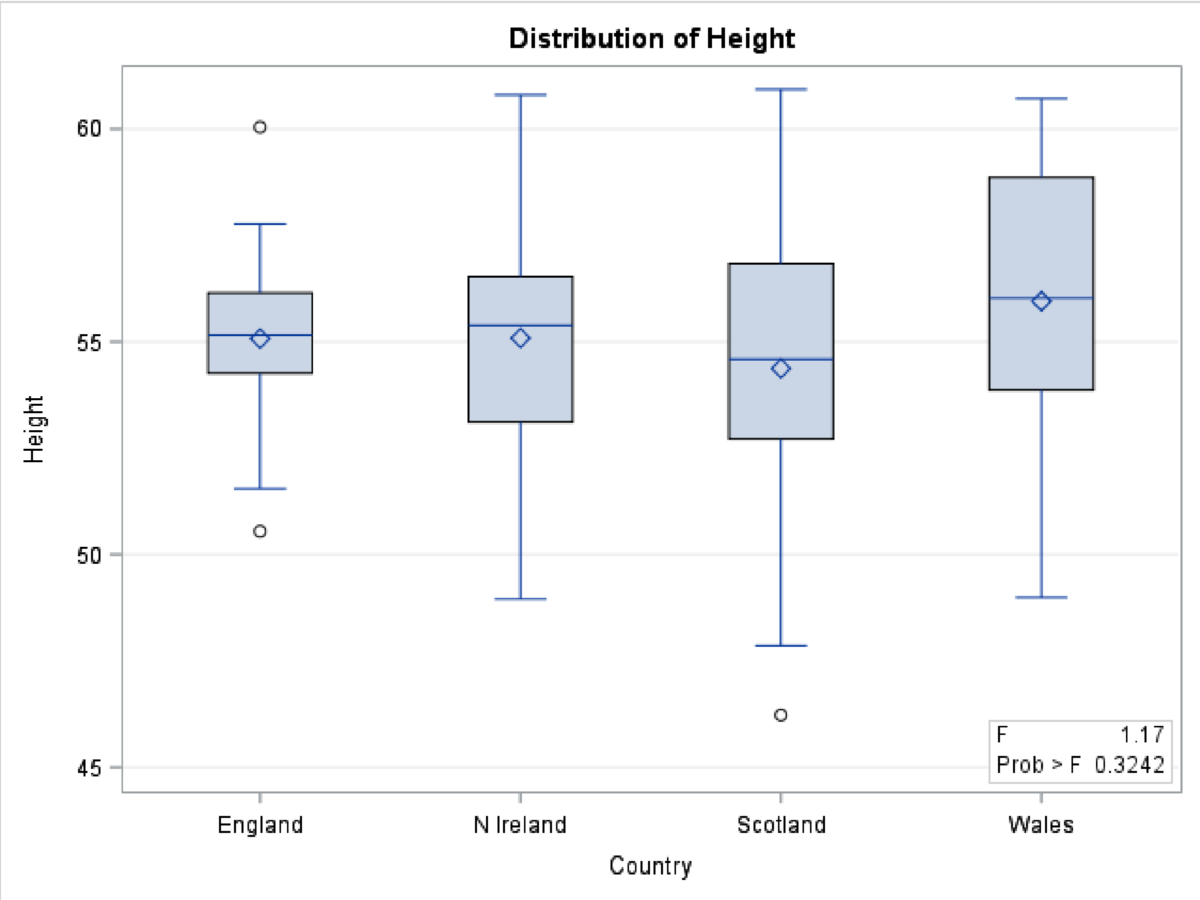
Examining Bartlett's test for Homogeneity of variances, we can see that a p-value of 0.0701 is not significant at the 5% level, therefore we do not reject the null hypothesis and we can conclude that there is no evidence to suggest that the variances are not equal.

| Bartlett's Test for Homogeneity of Height Variance | | | |
|---|---|---|---|
| **Source** | **DF** | **Chi-Square** | **Pr > ChiSq** |
| Country | 3 | 7.0580 | 0.0701 |

Results of the ANOVA show no significant difference at 5% (p-value =0.3242), therefore we do not reject the null hypothesis, and conclude that there is no evidence to suggest that the heights of 10 year olds differ by country.

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 3 | 31.3784751 | 10.4594917 | 1.17 | 0.3242 |
| Error | 96 | 856.0528874 | 8.9172176 | | |
| Corrected Total | 99 | 887.4313625 | | | |

The box and whisker plot further supports this conclusion as we can see the means are similar.



## 2. Counselling Therapy

In counselling and psychotherapy of subjects, the beneficial effects of pre-therapy training have been discussed. Four different approaches to pre-therapy training are (a) control (no training), (b) therapeutic reading (TR), which involves indirect learning, (c) vicarious therapy pre-training (VTP), which involves videotaped vicarious learning, and (d) group, role induction interview (RII), which involves direct learning. It is useful to compare the effectiveness of these approaches on the response variable, psychotherapeutic attraction, which is assigned a score between 0 and 40. Are there significant differences in the level effectiveness of approach? If so, which approach is the best and which is the worst?

It is stated in the question that the response variable is a score, therefore we need to perform a non-parametric one-way ANOVA, and interpret the output of the Kruskal-Wallis test.

The null and alternative hypotheses for this test are:

$H_0$: The m distributions are identical;

$H_1$: $N$ot all of the distributions are the same.

| Kruskal-Wallis Test | | |
|---|---|---|
| **Chi-Square** | **DF** | **Pr > ChiSq** |
| 22.9452 | 3 | <.0001 |

By examining the results of the Kruskal-Wallis (p-value < 0.0001) we can see that a significant result has been obtained at the 5% level, even at the 0.1% level. Therefore we reject the null hypothesis and conclude that there is a difference between the effectiveness of the approaches.

The next step is to determine where the significant differences are, therefore, post-hoc comparisons need to be performed.

With the result using the Pairwise Two-Sided Multiple Comparison Analysis as follows.

| Pairwise Two-Sided Multiple Comparison Analysis | | | |
|---|---|---|---|
| Dwass, Steel, Critchlow-Fligner Method | | | |
| Variable: Score | | | |
| **Group** | **Wilcoxon Z** | **DSCF Value** | **Pr > DSCF** |
| **Control vs. RII** | -2.7887 | 3.9438 | 0.0272 |
| **Control vs. TR** | -3.0559 | 4.3217 | 0.0120 |
| **Control vs. VTP** | -3.5836 | 5.0680 | 0.0019 |
| **RII vs. TR** | -1.7264 | 2.4414 | 0.3099 |
| **RII vs. VTP** | -3.1855 | 4.5049 | 0.0079 |
| **TR vs. VTP** | -2.0870 | 2.9514 | 0.1573 |

Using this table is can be seen that there is a significant difference between all pairs of groups except for TR against both RII and VTP. Therefore the Control group has significantly the lowest scores, which is what would be expected as no training has been provided and VTP along with TR have the highest scores (these two are not significantly different, hence you cannot determine which is higher, although VTP has the higher mean).

This result is further supported by examining the box plots. Note that there is one outlier in the data which should be investigated further.

## 3. Mercury Pollution

An ecologist wishes to assess the level of mercury contamination in five different rivers. To achieve this he catches thirty fish from each river, and measures their level of mercury concentration. Are there significant differences in the level of pollution? If so, which rivers are the worst and least effected?

Let $\mu_i$ be the mean mercury level in fish from river $i$.

Let p= Porthead, r= Reasbury, s= Swinford and w= Whitethorpe.

Then

$$H_0: \mu_p = \mu_r = \mu_s = \mu_w$$

$$H_1: at\ least\ one\ \mu_i\ not\ equal$$

We therefore need to now check that mercury level is normally distributed in fish from each river, the null and alternative hypotheses being that:

$$H_0: \text{samples } x_p,\ x_r, x_s, x_w \text{came from a normally distributed population}$$

$$H_1: at\ least\ one\ sample\ did\ not\ come\ from\ a\ normally\ distributed\ population$$

Looking at the p-values for all tests for normality, no significant difference is found. Therefore we do not reject the null hypothesis, so there is no evidence to suggest that the distribution is not a normally distributed.

The final assumption is that of equal variances, with the null and alternative hypothesis being:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ \dots\dots = \sigma_m^2$$

$$H_1: \sigma_i^2 \text{ are not all equal.}$$

Examining Bartlett's test for Homogeneity of variances, we can see that a p-value of 0.0701 is not significant at the 5% level, therefore we do not reject the null hypothesis and we can conclude that there is no evidence to suggest that the variances are not equal.

Bartlett's test provides a p-value of 0.6185, which is not significant at the 5% level, therefore we do not reject the null hypothesis (as stated previously) and we can conclude that the variances of each group are equal.

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 4 | 3990.968920 | 997.742230 | 56.15 | <.0001 |
| Error | 142 | 2523.145393 | 17.768630 | | |
| Corrected Total | 146 | 6514.114312 | | | |

Examining the result of the ANOVA we can see that a p-value of <0.0001 is significant at the 5% level, therefore we can reject the null hypothesis (as stated above) and conclude that there is a significant difference between the levels of pollution in the rivers.

The next step is to perform post-hoc comparisons to determine where the significant differences can be found, again the Bonferoni test will be used.

By examining the significances for each pair of rivers, it can be seen that a significant difference can be found between all rivers except for 'Reasbury and Swindford' and 'Kelee and Whitethorpe'.

This is further supported by the boxplot which shows that Reasbury and Swinford gave the highest readings (not significantly different from each other), so were worst effected and Porthead significantly gave the lowest readings, so was least effected.