Statistical Methods II

Quiz 02

Name: Date:

Welcome to the second quiz. I designed this quiz for you to see what information I need to explain again and to give you some practice for the first examination (February 24, 2011). You have five (5) minutes to answer these 10 questions. The total number of points is 10.

⋆ ⋆ ⋆

Distribution Review

1 (1 point)

Let us suppose $X \sim \mathcal{N}(\mu = 1, \sigma^2 = 3)$, $Y \sim \mathcal{N}(\mu = 3, \sigma^2 = 2)$. What is the distribution of $X + Y$?

Solution:

The sum of two Normal distributions is also a Normal distribution. The means add, as do the variances (as long as X and Y are independent). Thus,

$$X + Y \sim \mathcal{N}(\mu = 4, \sigma^2 = 5)$$

Hypothesis Testing for two samples

Let $x_1, x_2, \ldots, x_m$ be a sample from $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $y_1, y_2, \ldots, y_n$ be a sample from $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. In other words, the x's and the y's are samples from two populations that are Normally distributed, but we do not know the respective population means or population variances.

2 (1 point)

From the information given, what test should we use to compare the two means? A t-test or a Mann-Whitney test? Explain.

Solution:

We should use the t-test (`t.test`), as we know the populations are Normally distributed. If we did not know this, then we would have to test Normality of the sample to determine if we use the parametric test (the t-test) or the non-parametric test (Mann-Whitney).

3 (1 point)
From the information given, what test do we use to determine if the two variances are not equal, $\sigma_X^2 \neq \sigma_Y^2$? Give the name of the test or the `R` function.

> **Solution:**
>
> We have two Normally distributed populations. Thus, we would use the F-test of variances (`var.test`).

Hypothesis Testing for three samples
Let $x_1, x_2, \ldots, x_m$ be a sample from X, $y_1, y_2, \ldots, y_n$ be a sample from Y, and $z_1, z_2, \ldots, z_n$ be a sample from Z. In other words, the x's, y's, and z's are samples from three populations, where we do not know their distribution nor do we know their respective population means or population variances.

4 (1 point)
We would like to use ANOVA to test the equality of means. Give one reason why we do not want to perform 3 separate t-tests.

> **Solution:**
>
> If we perform three separate t-tests, we are performing multiple tests. As such, we will have to adjust our p-values to reflect the effects of these three tests. Furthermore, we are lazy and only want to perform one test, which our ANOVA procedure (`aov`) can do.

5 (1 point)
To use ANOVA, there are three assumptions that need to be checked. The first is independence. This is (for now) checked using logic of the experiment. What is a second? What do we do to test that assumption?

> **Solution:**
>
> The second assumption is that each population is Normally distributed. We can test this for each group individually by using the Kolmogorov-Smirnov test (`ks.test`) or the Shapiro-Wilk test (`shapiro.test`).

6 (1 point)
What is the third, and what do we do to test that assumption?

> **Solution:**
>
> The last assumption is equality of variances. As we have more than two groups, we cannot use the F-test of variances (`var.test`) without adjusting our p-values to reflect the effects of multiple tests.
> Thus, we will use the Bartlett test (`bartlett.test`).

True and False

Please write 'TRUE' in the space provided if the statement is true; 'FALSE', if the statement is not true.

7 (1 point)
The sum of seven independent Normal distributions is a Normal distribution.

7. _____

> **Solution:**
>
> TRUE. The sum of any number of Normal distributions is a Normal distribution. The expected values (means) will always add. The variances will simply add only if the distributions are independent. If the distributions are not independent, then covariances come into play, which we will discuss when we get to regression.

R Questions

What do the following commands do? Make sure you make it clear that you know what the function and its given parameters do.

8 (1 point)

```
wilcox.test(var1, var2)
```

> **Solution:**
>
> This test will perform the Mann-Whitney test, which is a non-parametric test comparing two sample means. Here, the two samples are `var1` and `var2`.

9 (1 point)
`shapiro.test(var1)`

> **Solution:**
>
> This statement will perform a test of the Normality of sample `var1`.

R Questions

I tested the IQ of three groups of patients. Group 1 consisted of patients who were not Autistic. Group 2 consisted of patients who were mildly Autistic. Group 3 consisted of patients who were severely Autistic. My research hypothesis was that IQ and Autism are not statistically related. To do this, I thought I should use an Analysis of Variance procedure. During the analysis, I typed in `bartlett.test(IQ Group, data=Autism)`. The following was the output.

```
Bartlett test of homogeneity of variances

data:  IQ by Group
Bartlett's K-squared = 1.8195, df = 2, p-value = 0.4026
```

10 (1 point)
What is the conclusion of this test?

> **Solution:**
>
> The Bartlett Test determines if the variances across the several groups are significantly different. As the p-value is greater than our usual $\alpha = 0.05$, we conclude that the group variances are not significantly different.