

Statistics Advance Assignment - 7

January 27, 2024

Q1. Write a Python function that takes in two arrays of data and calculates the F-value for a variance ratio test. The function should return the F-value and the corresponding p-value for the test.

```
[1]: import numpy as np
from scipy.stats import f
def var_ratio_test(arr1, arr2):
    var1 = np.var(arr1, ddof=1)
    var2 = np.var(arr2, ddof=1)
    f_value = var1 / var2
    df1 = len(arr1) - 1
    df2 = len(arr2) - 1
    p_value = f.cdf(f_value, df1, df2)
    return f_value, p_value
arr1 = [1, 2, 3, 4, 5]
arr2 = [6, 7, 8, 9, 10]
f_value, p_value = var_ratio_test(arr1, arr2)
print(f"F-value: {f_value:.2f}")
print(f"P-value: {p_value:.2f}")
```

F-value: 1.00

P-value: 0.50

Q2. Given a significance level of 0.05 and the degrees of freedom for the numerator and denominator of an F-distribution, write a Python function that returns the critical F-value for a two-tailed test.

```
[2]: from scipy.stats import f
def critical_f_value(dfn,dfd, alpha=0.05):
    f_crit = f.ppf(1-alpha/2, dfn, dfd)
    return f_crit
dfn = 3
dfd = 16
alpha = 0.05
f_crit = critical_f_value(dfn, dfd, alpha)
print(f"Critical F-value: {f_crit:.4f}")
```

Critical F-value: 4.0768

Q3. Write a Python program that generates random samples from two normal distributions with known variances and uses an F-test to determine if the variances are equal. The program should output the F-value, degrees of freedom, and p-value for the test.

```
[3]: import numpy as np
from scipy.stats import f
def f_test(sample1, sample2, var1, var2, alpha=0.05):
    s1_var = np.var(sample1, ddof=1)
    s2_var = np.var(sample2, ddof=1)
    f_value = s1_var / s2_var if s1_var > s2_var else s2_var / s1_var
    dfn = len(sample1) - 1 if s1_var > s2_var else len(sample2) - 1
    dfd = len(sample2) - 1 if s1_var > s2_var else len(sample1) - 1
    p_value = f.sf(f_value, dfn, dfd)
    if p_value < alpha:
        print("The null hypothesis that the variances are equal is rejected.")
    else:
        print("The null hypothesis that the variances are equal is not rejected.
        ↵")
    return f_value, dfn, dfd, p_value
np.random.seed(123)
sample1 = np.random.normal(10, 2, size=20)
sample2 = np.random.normal(10, 3, size=20)
f_value, dfn, dfd, p_value = f_test(sample1, sample2, var1=2**2, var2=3**2)
print(f"F-value: {f_value:.4f}")
print(f"Degrees of freedom (numerator): {dfn}")
print(f"Degrees of freedom (denominator): {dfd}")
print(f"P-value: {p_value:.4f}")
```

```
The null hypothesis that the variances are equal is not rejected.
F-value: 1.7070
Degrees of freedom (numerator): 19
Degrees of freedom (denominator): 19
P-value: 0.1264
```

Q4. The variances of two populations are known to be 10 and 15. A sample of 12 observations is taken from each population. Conduct an F-test at the 5% significance level to determine if the variances are significantly different.

```
[4]: import numpy as np
from scipy.stats import f
n1 = 12
n2 = 12
var1 = 10
var2 = 15
f_stat = var1 / var2
p_val = f.cdf(f_stat, n1 - 1, n2 - 1)
alpha = 0.05
```

```

if p_val < alpha / 2 or p_val > 1 - alpha / 2:
    print("The variances are significantly different.")
else:
    print("There is no significant difference between the variances.")

```

There is no significant difference between the variances.

Q5. A manufacturer claims that the variance of the diameter of a certain product is 0.005. A sample of 25 products is taken, and the sample variance is found to be 0.006. Conduct an F-test at the 1% significance level to determine if the claim is justified.

```

[5]: import numpy as np
from scipy.stats import f
n = 25
sample_var = 0.006
hypo_var = 0.005
f_stat = sample_var / hypo_var
p_val = f.cdf(f_stat, n - 1, np.inf)
alpha = 0.01
if p_val < alpha:
    print("The manufacturer's claim is not justified.")
else:
    print("The manufacturer's claim is justified.")

```

The manufacturer's claim is not justified.

Q6. Write a Python function that takes in the degrees of freedom for the numerator and denominator of an F-distribution and calculates the mean and variance of the distribution. The function should return the mean and variance as a tuple.

```

[7]: import numpy as np
from scipy.stats import f
def f_distribution_mean_var(dfn, dfd):
    mean = dfd / (dfd - 2)
    variance = (2 * dfd**2 * (dfn + dfd - 2)) / (dfn * (dfd - 2)**2 * (dfd - 4))
    return (mean, variance)
mean, variance = f_distribution_mean_var(5, 10)
print("Mean: ", mean)
print(f"Variance: {variance:.4f}")

```

Mean: 1.25

Variance: 1.3542

Q7. A random sample of 10 measurements is taken from a normal population with unknown variance. The sample variance is found to be 25. Another random sample of 15 measurements is taken from another normal population with unknown variance, and the sample variance is found to be 20. Conduct an F-test at the 10% significance level to determine if the variances are significantly different.

```
[8]: from scipy.stats import f
n1 = 10
n2 = 15
s1_sq = 25
s2_sq = 20
f_stat = s1_sq / s2_sq
p_val = f.cdf(f_stat, n1 - 1, n2 - 1)
alpha = 0.1
if p_val < alpha:
    print("The variances are significantly different.")
else:
    print("The variances are not significantly different.")
```

The variances are not significantly different.

Q8. The following data represent the waiting times in minutes at two different restaurants on a Saturday night:

Restaurant A: 24, 25, 28, 23, 22, 20, 27 **Restaurant B:** 31, 33, 35, 30, 32, 36.

Conduct an F-test at the 5% significance level to determine if the variances are significantly different.

```
[9]: from scipy.stats import f
a = [24, 25, 28, 23, 22, 20, 27]
b = [31, 33, 35, 30, 32, 36]
n1 = len(a)
n2 = len(b)
s1_sq = sum((x - sum(a)/n1)**2 for x in a) / (n1 - 1)
s2_sq = sum((x - sum(b)/n2)**2 for x in b) / (n2 - 1)
f_stat = s1_sq / s2_sq
p_val = f.cdf(f_stat, n1 - 1, n2 - 1)
alpha = 0.05
if p_val < alpha:
    print("The variances are significantly different.")
else:
    print("The variances are not significantly different.")
```

The variances are not significantly different.

Q9. The following data represent the test scores of two groups of students:

Group A: 80, 85, 90, 92, 87, 83 **Group B:** 75, 78, 82, 79, 81, 84.

Conduct an F-test at the 1% significance level to determine if the variances are significantly different.

```
[10]: from scipy.stats import f_oneway
group_a = [80, 85, 90, 92, 87, 83]
group_b = [75, 78, 82, 79, 81, 84]
f_statistic, p_value = f_oneway(group_a, group_b)
```

```
alpha = 0.01
if p_value < alpha:
    print("The variances are significantly different (reject null hypothesis)")
else:
    print("The variances are not significantly different (fail to reject null_
↪hypothesis)")
```

The variances are not significantly different (fail to reject null hypothesis)