

3 Univariate Statistics

3.1 Introduction

3.2 Empirical Distributions

The standard deviation describes the typical spread of the observations around the mean.

3.3 Examples of Empirical Distributions

Similarly, the largest bin center `vmax` equals the maximum value of the data `corg` minus half of the range of the data `corg` divided by `nbin`.

where `1.0` sets the relative bar width to 100%. Instead of using the above script, we can also use the functions `histcounts` and `histogram` to determine the bin centers, edges, and counts.

The MATLAB Help lists and explains various such methods for automatic binning. The function `histogram` also returns a histogram object `h`,

as well as the histogram plot. We can also use `histcounts` to determine `n` and `e` without plotting the histogram.

The functions `histogram` and `histcounts` provide numerous ways of binning the data, normalizing the data, and displaying the histogram.

The third measure in this context is the mode, which is approximately the midpoint of the interval with the highest frequency.

3.4 Theoretical Distributions

The probability distribution density $f(x)$ defines the probability that the variable has a value in the vicinity of x .

3.5 Examples of Theoretical Distributions

3.6 Hypothesis Testing

The *p-value* of a hypothesis test is the probability (under the null hypothesis) of observing values at least as extreme as observed for the test statistics (Fig. 3.14).

3.7 The t-Test

We can now compare the calculated `tcalc` value of 0.7279 with the critical

tcrit value. This can be accomplished using `tinv`, which yields the inverse of the t -distribution function with $na+nb-2$ degrees of freedom at a 5% significance level.

The function `tinv` yields the inverse of the t -distribution function with $na+nb-2$ degrees of freedom at a 5% significance level.

3.8 The F-Test

A 95% confidence interval is `[0.6429, 1.8018]`, which includes the theoretical (and hypothesized) ratio `var(corg1)/var(corg2)` of $1.2550^2/1.2097^2=1.0762$.

3.9 The χ^2 -Test

The function `chi2inv` computes the inverse of the χ^2 CDF with input arguments α and Φ .

3.10 The Kolmogorov–Smirnov Test

```
kscalC = max(cn_obs-normcdf(x,0,1))
```

The output variable `kscrit=0.1723` differs slightly from that in Tab. 3.1 since `kstest` uses a slightly more precise approximation for the critical value for sample sizes larger than 40 from Miller (1956).

3.11 The Mann–Whitney Test

3.12 The Ansari–Bradley Test

3.13 Distribution Fitting

The function `gmdistribution.fit` creates an object of the `gmdistribution` class (see Chap. 2.5 and MATLAB Help on object-oriented programming for details on objects and classes).

3.14 Error Analysis

After clearing the workspace and resetting the random number generator, we can generate measurement data by randomly sampling a population with a mean of 3.4 and a standard deviation of 1.5. This could mean that the true gold content of a rock is 3.4 grams per ton, but this value cannot be determined more precisely than ± 1.5 grams per ton due to measurement errors.

In our example, the first digit of the measurement uncertainty is 2, and we therefore use the first digit to the right of the decimal point (i.e., 7) as the rounding digit. The frequency distribution has a mean of 4.4 grams per ton and a standard deviation of 2.8 grams per ton. This wide range includes the true value of 3.4 grams per ton. If I were to take 12 repeated measurements of the gold content, the means of these measurements would have a standard deviation of approximately 0.8 grams per ton, which is the standard error of the mean.

As expected, the value is much smaller than the corresponding value for $n=12$.

Recommended Reading

Figure Captions