The University of Hong Kong – Department of Statistics and Actuarial Science – STAT2802 Statistical Models – Tutorial Problems

That is to say, they must be of a statistical nature, or as we shall say later on, they must be statistical hypotheses. –Neyman and Pearson

Problems 61-70 on Testing Statistical Hypotheses (STAT2802 Statistical Models Tutorial notes for the week of 19-NOV-2012)

Impossible consequence leads to impossible premise/hypothesis. This is a basic rule of logic. In search of patterns in our complex world, this rule of logic is constantly used to test the validity of various theories formed for the explanation of phenomena by identifying some consequences of the theory and then analyzing these consequences' possibility. When, often, the channel of information from the premise to the consequences is spoiled by randomness, we still hope that the randomness can be described reasonably well by a probability distribution. With such hope we adapt the original rule into one of statistical logic: improbable consequence leads to improbable premise. This is the logic of statistical hypothesis testing.

In the hands of a statistician are the data, modeled as an i.i.d. sample drawn from a population as represented by a probability distribution, up to some undetermined parameters of the distribution. In the mind of a statistician is the hypothesis, usually addressing the undetermined parameter(s). In the following, we restrict ourselves to testing those hypotheses about a parameter of the population distribution.

In the basic case, the statistician only wants to know whether the hypothesis is improbable, leading to its rejection. The process of a basic testing is outlined as the following.

1. [Choose a Model] Collect data: $\{Y_i\}_{i=1}^n$ & Decide on the population distribution: $\mathbb{P}^{Y_1}$ & State the Hypothesis: H_0 (usu. statement about a parameter)
2. [Choose a Test Statistic] Formulate a univariate statistic T on $\{Y_i\}_{i=1}^n$, preferably a sufficient statistic for H_0, with the hindsight that T's sampling distribution is either known or not too hard to deduce.
3. [Deduce the Sampling Distribution] Deduce $\mathbb{P}\{T|H_0\}$, the null distribution of T.
4. [Make a conclusion] based on either of the following approaches:
    a. [Compute the p-value] Calculate the probability of the set of those values of T that are more improbable than T's realization, i.e., the p-value of T's realization, which is invented as a measure of probability of a consequence of H_0.
    b. [Specify the Critical Region] Specify the critical region on the sample space of T with regard to $\mathbb{P}\{T|H_0\}$ and check whether the realization of T falls into the critical region.

The p-value is a handy quantitative index about the probability of a particular consequence of the hypothesis. Small p-value concludes improbability of the consequence, implying improbability of the hypothesis. Due to the presence of randomness, the above process could lead to wrong conclusions. There are two types of wrong conclusions: [**type-1 error**] rejecting the hypothesis when it is true; [**type-2 error**] accepting the hypothesis when it is false. Corresponding to this is the two types of correct conclusions: [**<u>one minus the significance</u> of the test**] accepting the hypothesis when it is true and [**power of the test**] rejecting the hypothesis when it is false.

The critical region approach is more general and flexible; however, often it is specified as the region of lowest probability under $\mathbb{P}(T|H_0)$ and becomes essentially equivalent to the p-value approach. Although the critical region is specified with some scientific knowledge, it is specified under $\mathbb{P}(T|H_0)$. The parameter's true value is unknown. It could be rather unexpected and thus different from what H_0 describes it. Therefore the specification of the critical region is essentially independent of the value of the parameter, though it specified under H_0. This leads to the creation of the *power function* that can be used to characterize the test and to compare different tests. The power function of a statistical hypothesis test is a function mapping each possible value of the parameter to the critical region's probability under that value of the parameter; the power function sends each point in the parameter space to a point on the unit interval. Test 1 is uniformly *more* powerful than Test 2 if the power of Test 1 is greater than the power of Test 2 at any possible value of the parameter. Test 1 is Uniformly *Most* Powerful among all tests if the power of test 1 is greatest among the powers of all tests at any possible value of the parameter.

An extension of the basic case is the case of judging between a *pair* of hypotheses; one of the pair is called the null hypothesis, the other alternative hypothesis. The legacy of the single hypothesis of the simple case above is inherited, usually, by the null hypothesis. Thus, the test's type-1 error/type-2 error/1 – significance/power are the type-1 error/type-2 error/1 – significance /power of the null hypothesis. The null hypothesis is usually more explicit, more specific, more attached to some scientific reasons, or simply more conservative; the alterative hypothesis could be, for example, simply the opposite of the null, or a less believed but still possible one from the scientific analysis. But this discrimination between the pair is soft and completely arises from practical concerns.

If a hypothesis completely pins down the population distribution, it is called a *simple* hypothesis; otherwise, when it only partially pins down the population distribution, it is called a *composite* hypothesis. When both null and alternative hypotheses are simple hypotheses, i.e., the test is in the form '$H_0: \theta = \theta_0$ v $H_1: \theta = \theta_1$', we are essentially comparing two possible parameter values and naturally we think of their likelihoods granted by the same sample of data. We will use the ratio of the two likelihoods as the test statistic. An important result, called the Neyman-Pearson Lemma, reveals that this test based on the likelihood ratio is most powerful among all simple-simple tests with the same Type-1 error.

Problems 61-70.

61. As a statistician, you want to determine whether a coin is fair. You tossed the coin n times and record the 0-1 outcomes as $\{Y_i\}_{i=1}^{n}$ and modeled the population distribution as $Y_1 \sim$Bernoulli(θ) where $\theta = \mathbb{P}(Y_1 = 1)$. You started with the belief that the coin *is* fair. In other words, your hypothesis is H_0: ________. Then you selected a sufficient statistic $T =$_______ and you know that T follows the distribution _______________. Specifically under H_0, T follows the distribution: _______________. Finally, you compute the p-value of T's realization as p-value=_______________________________________. In an alternative approach you specified the critical region with size 5% as $\mathbb{C}(T, H_0, 5\%) =$_______________________ and the critical region's power function is expressed as $p(\theta, \mathbb{C}) =$___________________.

62. As a scientist you collected the following data recording the 7 repetitions of measurements of the same quantity:

$$\{y_i\} = \{22.1, \quad 20.3, \quad 19.7, \quad 21.1, \quad 18.9, \quad 19.8, \quad 21.2\}.$$

You believe that Y_1 follows a normal distribution with mean μ and variance σ^2. You started with the hypothesis that the mean is 20, that is, H_0:_________. Then you selected a sufficient statistic $m =$___________ and you know that a simple algebraic transformation of m into a pivot $T(\mu) =$______________$\sim$ ___-distribution with ____ degrees of freedom. Specifically under H_0, the pivot becomes a statistic, $T(\mu =$ ___$) =$ ________ follows the ___-distribution with ____ degrees of freedom. Finally, you compute T's realization (under H_0) = ________ and its p-value=_______ which leads to the `[ acceptance | rejection ]` of H_0 at 5% significance level. Alternatively, you specified the critical region with size 5% as $\mathbb{C}(T, H_0, 5\%) =$_______________ which `[ excludes | includes ]` the realization of T, leading to the `[ acceptance | rejection ]` of H_0 at 5% significance level.

63. As a financial analyst you collected the following monthly log-return data of the past year on the same index:

$$\{y_i\} = \{0.41, \quad 0.21, \quad -0.04, \quad 0.03, \quad -0.54, \quad 0.19, \quad 0.31, \quad -0.08, \quad 0.35, \quad 0.04, \quad 0.68, \quad -0.20\}.$$

You assume that Y_1 follows a normal distribution with mean μ and variance σ^2, where the standard deviation σ is an important indicator of the volatility of the market. You started with the hypothesis that the standard deviation is 20%, that is, H_0: ________. Then you selected a pivot $T(\sigma) =$ ________ $\sim$ ___-distribution with ___ degrees of freedom. Specifically under H_0, the pivot becomes a statistic, $T(\sigma =$ ____$)$ = _________ follows the ___-distribution with ___ degrees of freedom. Finally, you compute T's realization (under H_0) = ________ and its p-value=_______ leads to the `[ acceptance | rejection ]` of H_0 at 5% significance level. Alternatively, you specified the critical region with size 5% as $\mathbb{C}(T, H_0, 5\%) =$_______________ which `[ excludes | includes ]` the realization of T, leading to the `[ acceptance | rejection ]` of H_0 at 5% significance level.

64. Two samples of data are collected: $\{x_i\}$ ={178.6, 185.3, 179.5, 175.1, 189.7} and $\{y_j\}$ ={160.8, 165.2, 168.3, 170.2, 177.5, 162.9, 164.5, 167.2, 178.1}. As a statistician, you modeled them as both normal: $X_i \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ and $Y_i \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$. You believe that the two populations have equal variance and you this belief: H_0:_________. You then select a pivot $T(\sigma_1^2, \sigma_2^2)$=___________ $\sim$ ___-distribution with ___ and ___ degrees of freedom. Specifically under H_0, $T(\sigma_1^2 = \sigma_2^2) =$__________ $\sim$ ___-distribution

with ___ and ___ degrees of freedom. Finally, you compute T's realization (under H_0) = ________ and its p-value=_______ which leads to the `[ acceptance | rejection ]` of H_0 at 5% significance level. Alternatively, you specified the critical region with size 5% as $\mathbb{C}(T, H_0, 5\%) =$_______________ which `[ excludes | includes ]` the realization of T, leading to the `[ acceptance | rejection ]` of H_0 at 5% significance level.

65. With the same data and model as in 61, construct a 95% confidence interval for θ.

66. With the same data and model as in 62, construct a 95% confidence interval for μ.

67. With the same data and model as in 63, construct a 95% confidence interval for σ.

68. With the same data and model as in 64, construct a 95% confidence interval for σ_1^2/σ_2^2 .

69. An examination was given to two classes consisting of 40 and 50 students, respectively. In the first class the mean grade was 74 with a standard deviation of 8, while in the second class the mean grade was 78 with a standard deviation of 7. Is there a significant difference between the performance of the two classes at a level of significance of (a) 0.05, (b) 0.01? What is the p-value of the test?

70. To test the hypothesis that a coin is fair, the following decision rules are adopted: Accept the hypothesis if the number of heads in a single sample of 100 tosses is between 40 and 60 inclusive and reject the hypothesis otherwise. (1) Find the probability of rejecting the hypothesis when it is actually correct; (2) Interpret graphically the decision rule and the result of part (1); (3) What conclusions would you draw if the sample of 100 tosses yielded 53 head? 60 head? (4) Could you be wrong in your conclusions to (c)? Explain; (5) What is the probability of accepting the hypothesis that the coin is fair when the actual probability of heads is $p = 0.7$? 0.6? 0.8? 0.9? (6) Construct the graph of the power function for the test.