

Solutions to Exam #2

1. Define the following terms:

(a) Significance level of a hypothesis test.

Answer: The significance level, α , of a hypothesis test is the largest probability that the test will reject the null hypothesis when the null hypothesis is true. \square

(b) Type II error

Answer: A Type II error in hypothesis test occurs when the test fails to reject the null hypothesis when it is in fact false. \square

(c) Power of a test

Answer: The power of a hypothesis test is the probability that the test will reject the null hypothesis when it is false. \square

(d) p -value

Answer: The p value of a hypothesis test is the probability that the test statistic will take on an observed value, or more extreme ones, under the assumption that the null hypothesis is true. \square

(e) Simple hypothesis

Answer: A simple hypothesis hypothesis is one that completely specifies certain distribution. \square

2. Let X_1, X_2, \dots, X_n be a set of random variables with mean μ and variance σ^2 , and let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ and $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. State the assumption(s) needed to make the following statements true.

(a) $\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$.

Answer: Assume that X_1, X_2, \dots, X_n are iid normal(μ, σ^2). \square

(b) $\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$.

Answer: Assume that X_1, X_2, \dots, X_n are iid normal(μ, σ^2). \square

3. In a simple genetic model, if both the mother and the father of a child have genotype Gg , then there is probability $1/4$ that the child will have genotype GG , probability $1/2$ of genotype Gg , and probability $1/4$ of genotype gg . In a random sample of 24 children having both parents with genotype Gg , it is found that 10 have genotype GG , 10 have genotype Gg , and 4 have genotype gg . Perform an appropriate test on the data to determine whether the data are consistent with the genetic model. Explain your reasoning.

Solution: Index the categories of genotypes GG , Gg and gg by 1, 2 and 3, respectively. Assuming that the given genetic model is true, we then get that the corresponding probabilities of observing children in each category are

$$H_o: \quad p_1 = \frac{1}{4}, \quad p_2 = \frac{1}{2}, \quad p_3 = \frac{1}{4},$$

respectively. This constitutes the null hypothesis that we would like to test against the alternative that the actual distribution of genotypes does not follow the one predicted by the model.

Assume that H_o is true, we then get that the expected counts in each category are then ones listed on the third column of Table 1. The

Category (i)	p_i	Expected Counts	Observed Counts
1	1/4	6	10
2	1/2	12	10
3	1/4	6	4

Table 1: Expected and predicted counts for genotypes GG , Gg and gg

observe counts are also listed in Table 1. It then follows that the value for the Pearson Chi-Square statistic, Q , for this test is

$$\hat{Q} = \frac{(10 - 6)^2}{6} + \frac{(10 - 12)^2}{12} + \frac{(4 - 6)^2}{6} = \frac{11}{3} \approx 3.67.$$

For large values of n , Q has an approximate $\chi^2(2)$ distribution, or an exponential(2) distribution. Using this information, we can then

approximate the p -value to be

$$p\text{-value} = P(Q \geq \hat{Q}) \approx \int_{\hat{Q}}^{\infty} \frac{1}{2} e^{-x/2} dx = e^{-\hat{Q}/2} \approx 0.16.$$

This p -value is bigger than the usual significance levels ($\alpha = 0.1, 0.05,$ or 0.01). Thus, we cannot reject the null hypothesis and conclude that the observations are consistent with the simple genetic model. \square

4. A test of $H_o: \lambda = 1$ against $H_1: \lambda > 1$ is based on a test statistic, T , with a $\text{Poisson}(\lambda)$ distribution. Suppose the observed value of the test statistic is $\hat{T} = 3$.

- (a) Compute the p -value for the test.

Solution: The p -value is the probability that the test statistic will take on the observed value, or more extremes ones, given that the null hypothesis is true. In this case, “more extreme” means $T > \hat{T}$ since the alternative hypothesis is phrased in the form $\lambda > 1$. Thus,

$$p\text{-value} = P(T \geq 3), \quad \text{given that } T \sim \text{Poisson}(1).$$

Thus,

$$\begin{aligned} p\text{-value} &= 1 - P(T \leq 2) \\ &= 1 - \sum_{k=0}^2 \frac{1}{k!} e^{-1} \\ &= 1 - e^{-1} \left(1 + 1 + \frac{1}{2} \right) \\ &= 1 - \frac{5}{2e} \\ &\approx 0.08. \end{aligned}$$

\square

- (b) Can H_o be rejected at the $\alpha = 0.05$ significance level?

Solution: Since p -value > 0.05 , we cannot reject H_o at the $\alpha = 0.05$ significance level. \square

5. Suppose that X_1, X_2, \dots, X_n is a random sample from a normal($\mu, 1$) distribution. We wish to test the hypothesis $H_o: \mu = \mu_o$ against the alternative $H_1: \mu \neq \mu_o$. A sample of size $n = 25$ is drawn and the following rejection region is set

$$R: |\bar{X}_n - \mu_o| > c,$$

for some critical value c .

- (a) Determine a value of c so that the significance level of the test is $\alpha = 0.05$.

Solution: The significance level, α , is the largest probability that the test will reject the null hypothesis given that it is true. Since the null hypothesis for this test is simple, we have that

$$\alpha = P(|\bar{X}_n - \mu_o| > c),$$

given that $\bar{X}_n \sim \text{normal}(\mu_o, 1/n)$, where $n = 25$. We then have that

$$\begin{aligned} \alpha &= P\left(\frac{|\bar{X}_n - \mu_o|}{1/5} > \frac{c}{1/5}\right) \\ &= P(|Z| > 5c), \end{aligned}$$

where $Z \sim \text{normal}(0, 1)$. Then,

$$\begin{aligned} \alpha &= 1 - P(|Z| \leq 5c) \\ &= 1 - P(-5c \leq Z \leq 5c) \\ &= 1 - P(-5c < Z \leq 5c) \\ &= 1 - (F_z(5c) - F_z(-5c)) \\ &= 2 - 2F_z(5c), \end{aligned}$$

from which we get that

$$F_z(5c) = 1 - \frac{\alpha}{2}.$$

Thus, if $\alpha = 0.05$ we have that

$$F_Z(5c) = 0.975.$$

We then obtain from a standard normal probabilities table that

$$5c \approx 1.96.$$

Consequently, $c \approx 0.39$

□

- (b) Find an expression in terms of standard normal probabilities for the power function of this test.

Solution: The power function of this test, $\gamma(\mu)$ is the probability that the test will reject the null hypothesis when $\mu \neq \mu_o$; that is,

$$\gamma(\mu) = P(|\bar{X}_n - \mu_o| > c) \quad \text{given that } \bar{X}_n \sim \text{normal}(\mu, 1/25),$$

for $\mu \neq \mu_o$. Thus, we can write $\gamma(\mu)$ as

$$\begin{aligned} \gamma(\mu) &= 1 - P(|\bar{X}_n - \mu_o| \leq c) \\ &= 1 - P(\mu_o - c < \bar{X}_n \leq \mu_o + c) \\ &= 1 - P(\mu_o - \mu - c < \bar{X}_n - \mu \leq \mu_o - \mu + c) \\ &= 1 - P\left(5(\mu_o - \mu) - 5c < \frac{\bar{X}_n - \mu}{1/5} \leq 5(\mu_o - \mu) + 5c\right) \\ &= 1 - P(5(\mu_o - \mu) - 5c < Z \leq 5(\mu_o - \mu) + 5c), \end{aligned}$$

where $Z \sim \text{normal}(0, 1)$. We therefore have that

$$\gamma(\mu) = 1 - (F_Z(5(\mu_o - \mu) + 5c) - F_Z(5(\mu_o - \mu) - 5c)), \quad (1)$$

where F_Z denotes the cdf of the standard normal distribution. □

- (c) If $\mu = \mu_o - \frac{1}{5}$, compute the probability of a Type II error.

Solution: The probability of a Type II error when $\mu = \mu_o - 1/5$ is

$$\beta(\mu_o - 1/5) = 1 - \gamma(\mu_o - 1/5).$$

Using the formula in equation (1) we obtain that

$$\beta(\mu_o - 1/5) = F_Z(1 + 5c) - F_Z(1 - 5c),$$

where $5c = 1.96$ by part (a) of this problem. We then have that

$$\begin{aligned}\beta(\mu_o - 1/5) &= F_Z(2.96) - F_Z(-0.96) \\ &= F_Z(2.96) - (1 - F_Z(0.96)) \\ &= F_Z(2.96) + F_Z(0.96) - 1.\end{aligned}$$

Using a standard normal probabilities table we get that

$$\beta(\mu_o - 1/5) \approx 0.9985 + 0.8315 - 1 = 0.83.$$

□