Introduction to Hypothesis Testing

Question 1

A manufacturer of 40-amp fuses wants to make sure that the mean amperage at which its fuses burn out is in fact 40.

If the mean amperage is higher than 40, the manufacturer might be liable for damage to an electrical system due to fuse malfunction.

To verify the amperage of the fuses, a sample of fuses is to be selected and inspected. A hypothesis test is performed on:

H0: μ = 40 versus **H1**: μ > 40

Which of the following statements are true for describing Type I and Type II errors? (Check all that apply.)

Answer:

- Type II: The mean amperage is too high, but the manufacturer concludes that it is acceptable which opens them up to liability for damaged electrical systems.
- Type I: The mean amperage is good, but the manufacturer concludes that it is too high and they launch an expensive and unnecessary recall process.

Question 2

Minor surgery on horses under field conditions requires a reliable short-term anesthetic producing good muscle relaxation, minimal cardiovascular and respiratory changes, and a quick, smooth recovery with minimal after effects so that horses can be left unattended.

A study reports that for a sample of n=73 horses to which ketamine is administered under certain conditions, the sample average lateral recumbency (lying-down) time was 18.86 min.

Lateral recumbency time is known to be normally distributed with a standard deviation of σ=8.6 min.

Do these data suggest that true average lateral recumbency time under these conditions is less than 20 min?

Test the hypotheses:

H0: μ = 20 versus **H1**: μ < 20

at a level of significance 0.10.

Give the appropriate critical value and conclusion.

Answer: -1.28, Fail to reject **H0**.

Question 3

The recommended daily dietary allowance for zinc among males older than age 50 years is 15 mg/day. A study reports the following summary data on intake for a sample of 12 males aged 65-74 years: $\bar{x} = 11.3$.

Assuming that the true distribution for zinc intake is normally distributed with variance 6.43, does the data indicate that the average daily zinc intake in the population of all males 65-74 falls below the recommended allowance?

Consider testing the hypotheses:

H0: $\mu = 15$ versus **H1**: $\mu < 15$ at a level of significance $\alpha = 0.05$.

Below what value does $\bar{x}$ need to be for us to reject **H0** in favor of **H1**? Give your answer to two decimal places.

Answer: 13.80

Correct. The value is $\mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$, which is $15 - 1.645\frac{\sqrt{6.43}}{\sqrt{12}} = 13.80$.

Question 4

The director of manufacturing for ACME Industries is interested in an online training program that can be used to train the firm's maintenance employees for machine-repair operations.

To evaluate the training method, the director of manufacturing has agreed to train 15 employees with the new approach. The director will approve the program as long as the mean training time stays below 50 days. However, the director firmly believes that it will take, on average, 50 or more days.

Assuming that training times are normally distributed with a standard deviation of $\sigma = 3.2$ days, what is the largest sample mean that could be observed that will make the company switch to the online program when testing the relevant hypotheses at level $\alpha = 0.01$?

Round your answer to 2 decimal places.

Answer: 48.08

Correct. The company will reject **H0**: $\mu \geq 50$, in favor of **H1**: $\mu < 50$, if $\bar{X} < 50 - 2.33\frac{3.2}{\sqrt{15}} \approx 48.08$.

Question 5

Which of the following statements is necessarily true in hypothesis testing.

Answer: The relative severity of one error over the other is problem dependent.

Constructing Tests

Question 1

A commonly prescribed drug for relieving anxiety is believed to be 60% effective. Experimental results with a new drug administered to a random sample of 100 adults who are suffering from anxiety showed that 68 experienced relief.

Is this sufficient evidence to conclude that the new drug is superior to the one commonly prescribed?

Carry out the relevant hypothesis test using a 0.03 level of significance. Give the appropriate critical value and decision.

Answer: 1.88, Fail to reject H_0. The new drug does not appear to be superior at the 0.03 level of significance.

Question 2

"Hardness" of a material in engineering is defined as the resistance to indentation. It is determined by measuring the permanent depth of the indentation.

An engineer measured the "Brinell hardness" of 6 pieces of ductile iron that were subcritically annealed. The resulting data were:

179, 156, 167, 183, 178, 165

Assuming that Brinell hardness is normally distributed with a standard deviation of $\sigma = 4.2$, consider testing:

H0: $\mu = 170$ versus **H1**: $\mu > 170$.

What is the P-Value for this test and what is the conclusion based on this data set at the 0.20 level of significance?

Answer: 0.2184, Fail to Reject H_0.

Question 3

Which option below best describes the following two plots?

|

|

|

|

|

|

Answer: Plot A indicates a heavy-tailed distribution and Plot B indicates a light-tailed distribution.

Question 4

A random sample of 10 chocolate energy bars from a certain company has, on average, 232 calories with a standard deviation of 15 calories.

Let μ be the true average calorie count for all energy bars of this brand.

Consider testing **H0**: $\mu = 220$ versus **H1**: $\mu \neq 220$.

Which of the following is a desirable shape of the power function for this test?

Answer:

|

|

|

|

|

Question 5

A manufacturer of ketchup uses a machine to automatically dispense ingredients into bottles that move along an assembly line. The machine is working properly when 14 ounces are dispensed.

Suppose that the amount dispensed is normally distributed with a true standard deviation of 0.08 ounces. A random sample of 20 bottles had a sample mean of 13.8 ounces per bottle.

Is there evidence that the machine should be stopped and recalibrated? Test the relevant hypotheses at the 0.05 level of significance.

Which of the following is true?

Answer: Reject that the machine is properly calibrated if the sample mean is below 13.96 ounces or above 14.04 ounces.

More Hypothesis Tests!

Question 1

A manufacturer of 40-amp fuses wants to make sure that the mean amperage at which its fuses burn out is in fact 40.

If the mean amperage is lower than 40, customers will complain because the fuses require replacement too often. If the mean amperage is higher than 40, the manufacturer might be liable for damage to an electrical system due to fuse malfunction.

To verify the amperage of the fuses, a sample of 10 fuses is selected and inspected. The average amperage at which fuses in the sample burned out was 38.8. The sample variance was 2.8.

Assuming burn out amperage is normally distributed, consider testing the hypotheses:

H0: $\mu = 40$ versus **H1**: $\mu \neq 40$.

Find the P-value for this test.

Answer: 0.0495

Question 2

The director of manufacturing for ACME Industries is interested in a computer-assisted training program that can be used to train the firm's maintenance employees for machine-repair operations. To evaluate the training method, the director of manufacturing has requested an estimate of the mean training time required with the computer-assisted program.

Suppose that management has agreed to train 15 employees with the new approach. The resulting sample mean training time is 53.87 days, and the resulting sample standard deviation is 6.82 days.

Assuming that training times are normally distributed, is there evidence in the data to conclude that the true mean training time is greater than 50?

Use $\alpha = 0.05$.

Give the appropriate test statistic, critical value, and conclusion.

Answer: The test statistic is approximately 2.2. The critical value is approximately 1.76. The data does suggest that the true mean training time is greater than 50 when using $\alpha = 0.05$.

Question 3

Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution with mean μ and variance σ^2.

Let S^2 be the sample variance.

What is the distribution of S^2?

Answer:

$$\Gamma\left(\frac{n-1}{2}, \frac{2\sigma^2}{n-1}\right)$$

Correct

Correct. $W := \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) = \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$, so $S^2 = \frac{\sigma^2 W}{(n-1)}$ has another gamma distribution where the second parameter is divided by $\frac{\sigma^2}{(n-1)}$.

Question 4

A national supermarket chain wants to redesign self-checkout lanes throughout the country. Two designs have been suggested. Data on customer checkout times, in minutes, have been collected at stores where the systems have already been installed. The results were as follows:

	System A	System B
Sample size	120	100
Sample mean	4.1	3.3
Sample standard deviation	2.1	1.5

Assuming normality of customer checkout times for both systems, consider testing:

H0: $\mu_A = \mu_B$ versus **H1**: $\mu_A > \mu_B$, where μ_A is the true mean for System A and μ_B is the true mean for System B.

Give the value of the appropriate test statistic to 2 decimal places.

Answer: 3.29

Correct

Correct! The test statistic is:

$$\frac{\bar{X}_A - \bar{X}_B - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{4.1 - 3.3 - 0}{\sqrt{\frac{2.1^2}{120} + \frac{1.5^2}{100}}} \approx 3.29$$

Question 5

When is Welch's t-test used to compare two population means, and what distributional assumptions are needed?

Answer:

Welch's t-test is used when at least one of the two sample sizes is small, the true variances for the populations are unknown, and we can't assume they are equal. We need to assume that the two populations are normally distributed.

Best Tests and Some General Skills

Question 1

Let X_1 be a sample of size 1 from the continuous uniform distribution over the interval from 0 to θ. This means that X_1 has pdf:

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & \text{otherwise} \end{cases}$$

Consider testing the hypotheses:

H0: $\theta = 1$ versus **H1**: $\theta < 1$

at a level of significance α, based on this single observation X_1.

Which of the following is a valid test?

Answer:

Reject H_0, in favor of H_1 if $X_1 < \alpha$.

Correct! The rejection rule has the form, "Reject H_0 if $X_1 < c$" and c is determined by solving $\alpha = P(X_1 < c; \theta = 1) = c$.

Question 2

Compute the following quantity:

$$\chi^2_{0.01,9} + t_{0.90,6}$$

Give your answer to two decimal places.

Answer: 20.23

R code: `qchisq(0.99, 9) + qt(0.10, 6)`

Question 3

Let $X_1, X_2, \ldots, X_n$ be a random sample from the $\Gamma(2, \beta)$ distribution.

Give the Neyman-Pearson ratio for finding the best test of:

H0: $\beta = 2$ versus **H1**: $\beta = 3$.

Answer:

$$\left(\frac{3}{2}\right)^n e^{\left(\frac{1}{2} - \frac{1}{3}\right) \sum_{i=1}^n X_i}$$

Question 4

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the continuous uniform distribution over the interval from 0 to 1, which has pdf:

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & \text{otherwise} \end{cases}$$

Find the pdf for the maximum value in the sample.

Answer:

$$f(x) = \begin{cases} n\theta^n x^{n-1}, & 0 < x < \theta \\ 0, & \text{otherwise} \end{cases}$$

Question 5

Let $X_1, X_2, \ldots, X_n$ be a random sample from the $N(\mu, \sigma^2)$ distribution where μ is known.

On which statistic is the best test of:

H0: $\sigma^2 = \sigma_0^2$ versus **H1**: $\sigma^2 > \sigma_0^2$

based?

Answer:

$$\sum_{i=1}^{n}(X_i - \mu)^2$$

Uniformly Most Powerful Tests and F-Tests

Question 1

A teacher believes that the standard deviation of scores for a particular Midterm that he gives every semester is 4 points. His current students claim that the standard deviation is more than 4 points.

Let σ^2 be the true variance for the Midterm scores.

A random sample of 10 Midterm scores from the current semester has an observed standard deviation of $s = 5.2$ points.

Assuming that test scores are normally distributed, consider testing the hypotheses:

H0: $\sigma^2 = 16$ versus **H1**: $\sigma^2 > 16$.

Give the P-value for the appropriate test, rounded to 3 decimal places.

Answer: 0.085

The P-value here is the probability that a $\chi^2(9)$ random variable is above $\frac{(9)(5.2^2)}{16}$. This can be computed in R using the command:

```
1 - pchisq(9 * 5.2 * 5.2 / 16, 9)
```

Question 2

Consider a random sample X_1, of size 1, from the continuous distribution with pdf:

$$f(x;\theta) = \begin{cases} 1 - \theta \cdot 2(x - \frac{1}{2}), & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

for some parameter $-1 < \theta < 1$.

Does a uniformly most powerful (UMP) test for:

H0: $\theta = 0$ versus **H1**: $\theta \neq 0$

exist?

Answer:

Yes, and the form of the test will be to reject H_0, in favor of H_1, if $X_1 \leq c$ for some constant c to be determined.

Question 3

A random sample of size 8 from the $N(\mu_1, \sigma_1^2)$ distribution has a sample variance of $s_1^2 = 13.2$.

An independent random sample of size 6 from a $N(\mu_1, \sigma_2^2)$ has a sample variance of $s_2^2 = 15.1$.

Is there evidence to suggest that $\sigma_1^2 \neq \sigma_2^2$?

Use $\alpha = 0.03$ level of significance.

Answer:

No. The test statistic is approximately 1.1439, and we would reject the null hypothesis that the variances are equal, in favor of the stated alternative, if it is greater than 6.4287.

Question 4

Let $X_1, X_2, \ldots, X_n$ be a random sample from the $N(\mu, 1)$ distribution.

Consider deriving the uniformly most powerful (UMP) test for:

H0: $\mu = \mu_0$ versus **H1**: $\mu > \mu_0$.

We begin by finding the best test for:

H0: $\mu = \mu_0$ versus **H1**: $\mu = \mu_1$,

where μ_1 is some fixed value that is greater than μ_0.

Which rejection rules will give you the UMP test?

Answer:

Test A: Reject H_0, in favor of H_1 if

$$\exp\left[-\frac{1}{2}\left(\sum_{i=1}^{n}(X_i-\mu_0)^2-\sum_{i=1}^{n}(X_i-\mu_1)^2\right)\right]\leq k$$

for some k to be determined.

Test B: Reject H_0, in favor of H_1 if

$$\sum_{i=1}^{n}(X_i-\mu_0)^2-\sum_{i=1}^{n}(X_i-\mu_1)^2\geq k$$

for some k to be determined.

Both tests A and B above.

Question 5

True or False. For a simple H_0 and simple H_1, the uniformly most powerful test of size α is the same as the best test.

Answer:

True

Correct! While a UMP test is usually associated with a composite alternate hypothesis, the best test here is UMP because it is best or most powerful for all of the possible values described by H_1. There is only one value!

Adventures in GLRTs

Question 1

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the continuous distribution with pdf:

$$f(x;\theta)=\begin{cases}\theta(1-x)^{\theta-1}, & 0<x<1\\ 0, & \text{otherwise}\end{cases}$$

for some parameter $\theta>0$.

Consider testing:

H0: $\theta=1$ versus **H1**: $\theta\neq 1$.

Which of the following statements are true? (Check all that apply.)

Answer

- A uniformly most powerful (UMP) test does not exist.

  Correct! The UMP test for **H0**: $\theta=1$ versus **H1**: $\theta<1$ has a different direction than the UMP test for **H0**: $\theta=1$ versus **H1**: $\theta>1$, so there is not one test that is UMP for θ on "both sides" of 1.

- Under H_0, $X_1, X_2, \ldots, X_n$ are iid from the uniform distribution over the interval $(0,1)$.

  Correct! When $\theta=1$, the pdf becomes $f(x;1)=1$ for $0<x<1$.

Question 2

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the continuous distribution with pdf:

$$f(x;\theta)=\begin{cases}\theta(1-x)^{\theta-1}, & 0<x<1\\ 0, & \text{otherwise}\end{cases}$$

for some parameter $\theta>0$.

Consider testing:

H0: $\theta=1$ versus **H1**: $\theta\neq 1$.

What is the maximum likelihood estimator needed for computing the denominator of the generalized likelihood ratio (GLR)?

Answer:

$$\hat{\theta}=\frac{-n}{\sum_{i=1}^{n}\ln(1-X_i)}$$

Question 3

True or False.

A generalized likelihood ratio test will always be less powerful than a UMP test.

Answer: False

Correct

The GLRT and UMP tests are often the same test and therefore would have the exact same power function!

Question 4

True or False.

When computing a generalized likelihood ratio (GLR) for some H_0 versus some H_1, the restricted MLE is always less than or equal to the unrestricted MLE.

Answer: False

Correct

The restricted MLE, $\hat{\theta}_0$ can be smaller, larger, or equal to the unrestricted MLE $\hat{\theta}$. However, it is true that $L(\hat{\theta}_0) \leq L(\hat{\theta})$.

Question 5

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from the $\Gamma(2, \beta)$ distribution. Suppose that n is "large".

Consider testing:

H0: $\beta = 1$ versus **H1**: $\beta \neq 1$

using an approximate large sample generalized likelihood ratio test.

Suppose that the generalized likelihood ratio $\lambda(\vec{X})$ is observed to be $\lambda(\vec{x}) = 0.67$.

Which of the following is an approximate large sample test using $\alpha = 0.05$?

Answer:

The rule is to reject H_0, in favor of H_1 if $-2 \ln \lambda(\vec{X}) > 3.841$. For the given $\lambda(\vec{x})$, we fail to reject H_0.

Correct

Correct. $-2 \ln \lambda(\vec{x}) \approx 0.8$ and 3.841 is the appropriate χ^2 critical value. Since 0.8 is not greater than 3.841, we fail to reject H_0.