

Homework_1

Mohammad Imtiaz Nur

2/08/2021

ID: 1878074

Solution Task 1:

Statement (1) is incorrect. Given the meedian of total distribution is 100, so half of the scores can't be less than 95.

Statement (2) is also incorrec. As the upper quartile(Q3) = 120 and the lower quartile(Q1) = 70 and the middle 50% of the scores must be between Q1(70) and Q3(120), but not between 100 and 120.

Statement (3) is correct. Given that the upper quartile value is 120, the one-quarter of scores have values greater than 120.

Statement (4) is incorrect. As the mode is given as 75, it should be the most common score.

Solution Task 2:

(a)

For two independent event X and Y which are normally distributed; $P(Y > 12 \text{ and } X > 4) = P(Y > 12) \cdot P(X > 4)$

Calculating using R:

```
Y<- 1-pnorm(12, 10, sd = 2)
Y
```

```
## [1] 0.1586553
```

```
X<- 1-pnorm(4, 5, sd = 5)
X
```

```
## [1] 0.5792597
```

```
Y*X
```

```
## [1] 0.0919026
```

Therefore, $P(Y > 12) \cdot P(X > 4) = 0.1586553 \cdot 0.5792597 = 0.0919026$

(b)

For OR operation, Y and X both can be either mutually exclusive or not. Here I provided both calculations.

mutually exclusive:

$$P(Y > 12 \text{ or } X > 4) = P(Y > 12) + P(X > 4)$$

```
Y<- 1-pnorm(12, 10, sd = 2)
Y
```

```
## [1] 0.1586553
```

```
X<- 1-pnorm(4, 5, sd = 5)
X
```

```
## [1] 0.5792597
```

```
Y+X
```

```
## [1] 0.737915
```

$$\text{So, } P(Y > 12) + P(X > 4) = 0.1586553 + 0.5792597 = 0.737915$$

not mutually exclusive:

$$P(Y > 12 \text{ or } X > 4) = P(Y > 12) + P(X > 4) - P(Y > 12 \text{ and } X > 4)$$

```
Y+X- (X*Y)
```

```
## [1] 0.6460124
```

$$P(Y > 12) + P(X > 4) - P(Y > 12 \text{ and } X > 4) = 0.6460124$$

(c)

$$P(Y > 10 \text{ and } X < 5) = P(Y > 10) \cdot P(X < 5)$$

```
Y<- 1-pnorm(10, 10, sd = 2)
Y
```

```
## [1] 0.5
```

```
X<- pnorm(5, 5, sd = 5)
X
```

```
## [1] 0.5
```

Y * X

[1] 0.25

So, $P(Y > 10) \cdot P(X < 5) = 0.5 * 0.5 = 0.25$

Solution Task 3:

Given, $n = 200$, $p = 0.015$

Therefore, $\mu = np = 0.015 * 200 = 3$

From poisson's distribution, probability of getting no occurrences of the disease in a random sample of 200,

$$\begin{aligned} P(Y = 0) &= \frac{3^0 e^{-3}}{0!} \\ &= 0.0498 \end{aligned}$$

Solution Task 4:

###(a) You are deciding whether you should take an umbrella to work. H_0 = should take an umbrella

H_1 = shouldn't take an umbrella

Type-I error: I didn't take the umbrella and it rained.

Type-II error: I brought the umbrella but it didn't rain.

Here, TypeII error can be ignorable as it won't be as bad as TypeI error considering I will get wet if I don't take an umbrella in work. Hence as this case isn't too severe we can keep significance level (α) at an standard level of 0.05.

###(b) You are planning a proficiency testing procedure to determine whether some employees should be fired.

H_0 = employees shouldn't be fired

H_1 = employees should be fired

Type-I error: A proficient employee is fired.

Type-II error: An employee isn't fired even not being proficient.

Here, TypeI error will have serious consequences in growth of company as well as it will effect other proficient employees working in the company considering they might not evaluated properly. So, a lower value of α (less than 0.05) should be used in this case.

###(c) Same as part (b) except you want to determine whether some employees deserve a special merit raise.

H_0 = don't deserve a special merit raise

H_1 = deserve a special merit raise

Type-I error: An employee got merit raise even not being proficient.

Type-IT error: A proficient employee didn't get merit raise.

Here, TypeI error will have serious consequences among other employees and it will also make the test questionable. So, a lower value of α (less than 0.05) should be used in this case.

###(d) A cigarette manufacturer is conducting a test of nicotine content in order to justify a new advertising claim. H_0 = Conducting a test of nicotine content.

H_1 = Don't conduct the test of nicotine content.

Type-I error: A new advertising claim was made but the test to justify the claim is not conducted.

Type-II error: The new advertising claim was not made but the test was conducted for justification.

For this case, alpha can be considered as 0.05 as the TypeI error doesn't involve any significant cost for the test.

###(e) You are considering the procedure to decide guilt or innocence in a court of law.

H_0 = suspect is innocent

H_1 = suspect is guilty

Type-I error: An innocent got punished.

Type-II error: Guilty person set to be free.

Here, the TypeI error is very much severe as an innocent person will have to face jail for a crime he didn't do. In this case, α should be 0.001.

###(f) You are wondering whether you should buy a new battery for your calculator before the next statistics test.

H_0 = should buy a battery

H_1 = shouldn't buy a battery

Type-I error: didn't buy a battery and it failed.

Type-II error: I bought a battery but it wasn't needed.

In a statistics test, calculator is a vital thing without which anyone might be failed. So, for avoiding TypeI error we should consider α as 0.001.

###(g) As a university administrator you are considering a policy to restrict student driving in order scholastic achievement.

H_0 = Don't impose restriction on student driving

H_1 = Impose the restriction on student driving

Type-I error: A new policy to restrict student driving is imposed but scholastic achievement wasn't achieved

Type-II error: Policy not imposed and achieved scholastic achievement

To impose a new policy, a lot of administrative process including significant workforce and costs is necessary. For TypeI error, these effort will go in vain. So considering significance level of 0.05 for this case is standard.

Solution Task 5:

Here, H_0 = The new program doesn't reduce the time of processing.

To verify our null hypothesis with given means and standard deviation of both the distributions as parameters, we can write hypothesis test as following:

```
# m1, m2: the sample means
# s1, s2: the sample standard deviations
# n1, n2: the same sizes
# m0: the null value for the difference in means to be tested for. Default is 0.
# equal.variance: whether or not to assume equal variance. Default is FALSE.
```

```

t.test <- function(m1,m2,s1,s2,n1,n2,m0=0,equal.variance=FALSE)
{
  if( equal.variance==FALSE )
  {
    se <- sqrt( (s1^2/n1) + (s2^2/n2) )
    # welch-satterthwaite df
    df <- ( (s1^2/n1 + s2^2/n2)^2 )/( (s1^2/n1)^2/(n1-1) + (s2^2/n2)^2/(n2-1) )
  } else
  {
    # pooled standard deviation, scaled by the sample sizes
    se <- sqrt( (1/n1 + 1/n2) * ((n1-1)*s1^2 + (n2-1)*s2^2)/(n1+n2-2) )
    df <- n1+n2-2
  }
  t <- (m1-m2-m0)/se
  dat <- c(m1-m2, se, t, 2*pt(-abs(t),df), pt(-abs(t),df))
  names(dat) <- c("Difference of means", "Std Error", "t", "p-value (two-tailed)", "p-value (one-tailed)")
  return(dat)
}

computer.t.test<- t.test(12.3, 10.9, 3.5, 3.5, 100, 100)
computer.t.test

```

```

## Difference of means          Std Error          t
##      1.400000000          0.494974747      2.828427125
## p-value (two-tailed) p-value (one-tailed)
##      0.005158849          0.002579424

```

Here both p values (2-tail: 0.005158849, 1-tail: 0.002579424) is less than 0.05.

Hence, our null hypothesis can be rejected and therefore the new program can process the billing faster.