This is the tenth homework assignment. Students should tick in TUWEL problems they have solved and upload their detailed solutions by **20:00** on Monday **January 9th, 2023**.

---

(1) **One-sample test for proportions**
You are at a fair and a lottery booth showman claims that every second lottery ticket is a win. You observe the hustle and bustle around the lottery booth for a while and count that from 58 tickets sold 17 won. Does this observation let you doubt the claim?

(a) Calculate the p-value in the context of the (two-sided) one-sample test for proportions.

(b) What do you answer the showman?

(2) **One-sample test for proportions (without R)**
In the context of the one-sample situation for proportions let the observed relative frequency be $h = \frac{1}{2}$. Let the null hypothesis be $H_0 : p = 0.4$ and further let the (approximate) test be two-sided. Answer the following questions only using the table below, which shows the α-quantiles q_α of the $\mathcal{N}(2, 1)$-distribution.

α	0.01	0.05	0.1	0.2	0.5
q_α	-0.32	0.36	0.72	1.16	?

(a) What is the value of the question mark in the table?

(b) For $n = 49$ the null hypothesis is rejected on the 10%-level?

(c) For $n = 144$ the null hypothesis is rejected on the 3%-level?

(3) **Which statement is correct?**
In the situation of the two-sample test for proportions, the null hypothesis that the population proportions are equal was not rejected at the 3%-level. (Let the sample sizes be large in the sense that the normal approximation is accurate). Comment on the following statements.

(a) The two observed relative frequencies are equal.

(b) The observed relative frequencies are equal if and only if the absolute frequencies are equal.

(c) The test statistic was larger than the 95%-quantile of the standard normal distribution.

(d) If we had performed a right-sided test, then (c) would be true.

(e) The 99%-confidence interval overlaps zero.

(f) The null hypothesis is correct with probability 97%.

(g) If the null hypothesis holds true, then a type-I error was not made.

(4) **Estimator for the variance**

Let $Y_1, \dots, Y_n$ be independent and identically distributed random variables with variance σ^2. The canonical estimator for σ^2 is the empirical variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2,$$

which we showed in a previous homework was unbiased for σ^2. Let $Y_1 \sim Ber(p)$ and H be the relative frequency of successes. In this special case we estimated σ^2 via $H(1-H)$. Is this estimator 'far away' from S^2? More precisely, express S^2 via $H(1-H)$.

(5) **Simulation of coverage probability**

Simulate the coverage probability of the one-sample confidence interval for frequencies - does the confidence interval deliver what it promises? Let $Y_1, \dots, Y_n$ be i.i.d. random variables with $Y_1 \sim bern(p)$ and $p \in (0,1)$. Approximate in 10000 simulations the coverage probability of the 95%-confidence interval, i.e., simulate the proportion of coverage events of the parameter p. For that let

(a) $n = 45$ and $p = 4/9$

(b) $n = 10$ and $p = 1/10$

(c) Visualize the simulated relative frequencies from (a) and (b) in two histograms and comment on the simulated coverage probabilities.

Hint: `R`-command `rbinom` and for example `ifelse()`.

(1) **One-sample test for proportions**

You are at a fair and a lottery booth showman claims that every second lottery ticket is a win. You observe the hustle and bustle around the lottery booth for a while and count that from 58 tickets sold 17 won. Does this observation let you doubt the claim?

(a) Calculate the p-value in the context of the (two-sided) one-sample test for proportions.

(b) What do you answer the showman?

$H_0: p = 0.5$

$H_a: p \neq 0.5$

$n = 58$

$p_0 = 0.5$

$q_0 = 0.5$

$h = \frac{17}{58} = 0{,}293$

$$se_h = \sqrt{\frac{h(1-h)}{n}} = 0{,}0597$$

$$z = \frac{h - p_0}{se_H} = \frac{0{,}293 - 0{,}5}{0{,}0597} = -3{,}467$$

$$\underbrace{|h - p_0|}_{0{,}207} \approx 3{,}467 \cdot \underbrace{se_h}_{0{,}0597}$$

→ this is pretty far

typical deviation of H is $1 \cdot SE_H$

$$p \text{ val} = 0{,}0003 \cdot 2 = 0{,}0006$$

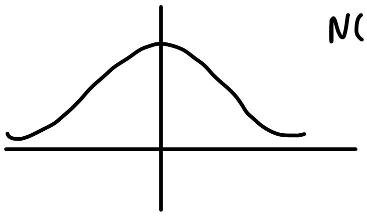
b) The chances are very low. The chance to win with 50% are 0,06% → will not happen

(2) **One-sample test for proportions (without R)**
In the context of the one-sample situation for proportions let the observed relative frequency be $h = \frac{1}{2}$. Let the null hypothesis be $H_0 : p = 0.4$ and further let the (approximate) test be two-sided. Answer the following questions only using the table below, which shows the α-quantiles q_α of the $\mathcal{N}(2, 1)$-distribution.

α	0.01	0.05	0.1	0.2	0.5
q_α	-0.32	0.36	0.72	1.16	?

(a) What is the value of the question mark in the table?
(b) For $n = 49$ the null hypothesis is rejected on the 10%-level?
(c) For $n = 144$ the null hypothesis is rejected on the 3%-level?

a)



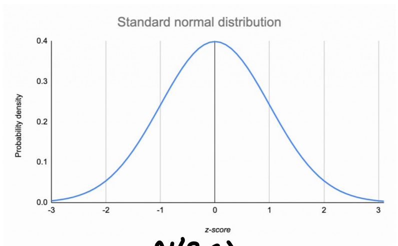
$N(2,1)$

$\rightarrow$ symmetric

& mean of 2 $\Rightarrow \alpha_{0.5} = 2$

b) $n = 49$

we need to standardize



$N(0,1)$

α	0.01	0.05	0.1	0.2	0.5
q_α	-0.32	0.36	0.72	1.16	?

α	0.02	0.1	0.2	0.4	1] $\alpha = 100\%$
$q_{\alpha/2}$	-2,32	-1,64	-1,28	-0,84	0

$R = (-\infty, -1{,}64] \cup [1{,}64, \infty)$

$h = \frac{1}{2}$ $n = 49$

$se_h = \sqrt{\frac{h(1-h)}{n}} \approx 0{,}714$

$z = \frac{h - p_0}{se_h} = \frac{0{,}5 - 0{,}4}{0{,}714} \approx 1{,}4$

$z \notin R \rightsquigarrow$ FTR H_0

c) $h = \frac{1}{2}$ $n = 144$

$se_h = \sqrt{\frac{h(1-h)}{n}} \approx 0{,}0417$

$z = \frac{h - p_0}{se_h} = \frac{0{,}5 - 0{,}4}{0{,}0417} \approx 2{,}4$

2% level

we know

α	0.02
$q_{\alpha/2}$	-2,32

so $\frac{0.02}{2{,}32}$

$2{,}4 > 2.32$ so we already reject for 2% level
$\Rightarrow$ also for 3% level

(3) **Which statement is correct?**
In the situation of the two-sample test for proportions, the null hypothesis that the population proportions are equal was not rejected at the 3%-level. (Let the sample sizes be large in the sense that the normal approximation is accurate). Comment on the following statements.

(a) The two observed relative frequencies are equal. could be we dont know

(b) The observed relative frequencies are equal if and only if the absolute frequencies are equal. False

(c) The test statistic was larger than the 95%-quantile of the standard normal distribution. can be left sided / right sided / or two sided → dont know

(d) If we had performed a right-sided test, then (c) would be true. False

(e) The 99%-confidence interval overlaps zero. True

(f) The null hypothesis is correct with probability 97%. I cant say about correctness

(g) If the null hypothesis holds true, then a type-I error was not made. True