Publish Preview Edit ⋮

This quiz is unpublished
Only teachers can see the quiz until it is published.

Practice version Exam Business Statistics 2020/06/05

ⓘ This is a preview of the draft version of the quiz

Quiz type	Graded quiz
Points	32
Assignment group	Imported Assignments
Shuffle answers	No
Time limit	120 Minutes
Multiple attempts	Yes
Score to keep	Latest
Attempts	Unlimited
View responses	Always
Show correct answers	Immediately
One question at a time	No
Require Respondus LockDown Browser	No
Required to view quiz results	No
Webcam Required	No

Due	For	Available from	Until
-	Everyone	-	-

Preview

Score for this attempt: **28** out of 32 *
Submitted 15 Jun at 22:54
This attempt took 52 minutes.

Question 1

1 / 1 pts

Welcome to the BSTAT written exam.

You are looking at the version for normal time students (so 120 + 5 minutes for the selfie). If you are entitled to **extra time**, close the current quiz now and go to the extra time exam quiz.

For emergency questions during the exam (the real exam that is), you can mail me your emergency question (c.s.bos@vu.nl) using the subject line starting with "BSTAT: ". I will check mail regularly.

Before proceeding, choose one of the statements below that applies to you.

- I will use the help of people from the internet during this exam.

Correct!

- I will make this exam by myself without any outside help for the entire duration of the exam.
- I will use the help of some friends during this exam.

Question 2

1 / 1 pts

Let X be a normal random variable $N\left(\mu = -200, \sigma^2 = 1024\right)$.

Compute the probability $P(X \geq -156)$ in 4 decimal positions.

(Use 4 decimals behind the decimal point in your answer. Also, use a decimal point, *not* a comma and only round your *final* answer and not any intermediate results.)

Correct!

0.0846

rrect Answers

0.084566 (with margin: 0.002825)

Question 3

1 / 1 pts

Let X_1 and X_2 be two independent normal random variables, both $N(\mu=110, \sigma^2=625)$ distributed. Let $\overline{X}=0.5(X_1+X_2)$. Compute $P(\overline{X} \geq 95)$ in 4 decimal positions.

(Use 4 decimals behind the decimal point in your answer. Also, use a decimal point, *not* a comma and only round your *final* answer and not any intermediate results.)

Correct!

0.8019

rrect Answers

0.801928 (with margin: 0.004972)

Question 4

1 / 1 pts

Let Y be a binomial random variable with a $\mathrm{Binomial}(n=8, \pi=0.8)$distribution. Compute the probability $P(Y \geq 5)$ in 4 decimal positions.

(Use 4 decimals behind the decimal point in your answer. Also, use a decimal point, *not* a comma and only round your *final* answer and not any intermediate results.)

Correct!

0.9437

rrect Answers

0.943718 (with margin: 0.00048)

Question 5

1 / 1 pts

You sell high end luxury goods. The probability that you close at least one sale is $\pi = 60\%$ per day. Assume that sales are independent across days and that you are open 5 days a week. Compute the probability of having no sales in a particular week in 4 decimal places.

(Use 4 decimals behind the decimal point in your answer. Also, use a decimal point, *not* a comma and only round your *final* answer and not any intermediate results.)

Correct!

0.0102

rrect Answers

0.01024 (with margin: 0.0002)

Question 6

0 / 1 pts

In a sample of size 15 from a normally $N(\mu, \sigma^2)$ distributed population, we find a sample mean of 188.0 and a sample standard deviation of 24.5. Give the *LOWER* bound of the

95% confidence interval for μ in 3 decimals.

(Use 3 decimals behind the decimal point in your answer. Also, use a decimal point, *not* a comma and only round your *final* answer and not any intermediate results.)

ı Answered 175.602

ɾrect Answers 174.432352 (with margin: 0.006326)

Question 7

1 / 1 pts

Given is a sample of size 24 of observations of the two variables X and Y, we computed a sample correlation of $r = -0.47$. We test the hypothesis $H_0 : \rho = 0$ against a 2-sided alternative using our standard test statistic for this purpose. Compute the value of this test statistic in 3 decimals.

(Use 3 decimals behind the decimal point in your answer. Also, use a decimal point, *not* a comma and only round your *final* answer and not any intermediate results.)

Correct! -2.4975

ɾrect Answers -2.497541 (with margin: 0.002)

Question 8

1 / 1 pts

For a quality control study, you measure the temperature of an item right after production in two different facilities. You are interested in testing whether the mean temperature at facility 1 (sample 1) is significantly lower than at facility 2 (sample 2). What is your conclusion at $\alpha = 0.10$ and why? You can use the following tables.

```
          n    mean     sd
Sample1  14  86.606  4.131
Sample2  14  90.636  4.922
```

```
Levene's Test for Homogeneity of Variance
 F value  P(>F)
   0.049  0.826
```

```
Two sample t-test
t = -2.346, df = 26
(Null hypothesis for this test: true difference in means is
0)
```

- [ ] Reject H_0 as the Levene test is not significant
- [ ] Do not reject H_0 as the Levene test is not significant
- [ ] Reject H_0 as $\mu_1 - \mu_2 < 0$
- [ ] Do not reject H_0 as $\mu_1 - \mu_2 < 0$

Correct!

- [x] Reject H_0 as the test t-statistic -2.346 is in the critical region
- [ ] Do not reject H_0 as the test t-statistic -2.346 is in the critical region
- [ ] Reject H_0 as the test t-statistic -2.346 is *not* in the critical region
- [ ]

Do not reject H_0 as the test t-statistic -2.346 is *not* in the critical region

Question 9

0 / 1 pts

For a quality control study, you measure the temperature of an item right after production in two different facilities. You want to test $H_0 : \mu_1 - \mu_2 = 9.3$ at $\alpha = 0.05$. Compute the value of the usual test statistic to test this hypothesis, using 3 decimal places.

(Use 3 decimals behind the decimal point in your answer. Also, use a decimal point, *not* a comma and only round your *final* answer and not any intermediate results.)

```
         n    mean     sd
Sample1  9   78.391   4.503
Sample2  9   74.485   2.897
```

```
Levene's Test for Homogeneity of Variance
 F value  P(>F)
   1.012  0.329
```

```
Two sample t-test
t = 2.188, df = 16
(Null hypothesis for this test: true difference in means is 0)
```

ı Answered

-2.9799

ırrect Answers

-3.022387 (with margin: 0.002)

-3.022387 (with margin: 0.002)

Question 10

1 / 1 pts

For a quality control study, you measure the temperature of an item right after production in two different facilities. You want to 'prove' that μ_1 is greater than μ_2. Give the appropriate NULL hypothesis.

- $H_0 : \mu_1 \geq \mu_2$
- **Correct!** $H_0 : \mu_1 \leq \mu_2$
- $H_0 : \mu_1 = \mu_2$
- None of the above

Question 11

1 / 1 pts

For a quality control study, you measure the temperature of an item right after production in two different facilities. Summary statistics are in the table below. You want to test $H_0 : \mu_1 \leq 42$ versus $H_1 : \mu_1 > 42$. Compute the CRITICAL value of the usual test statistic using 3 decimal places and $\alpha = 0.01$.

(Use 3 decimals behind the decimal point in your answer. Also, use a decimal point, *not* a comma and only round your *final* answer and not any intermediate results.)

```
         n    mean     sd
Sample1  14  43.887  7.365
Sample2  14  48.041  6.686
```

Correct!

2.6503

rrect Answers 2.650309 (with margin: 0.002)

Question 12

1 / 1 pts

A new test has been developed to detect a disease. It is known that 2% of the population has the disease. If a person has the disease, the test will be positive (detect that the person has the disease) with probability 97%. Conversely, if a person is healthy, the test will be falsely positive with a probability of 2% (and so will indicate in 98% of those cases that the tested person indeed does not have the disease).

Compute the probability that a randomly chosen person will test positive for the disease using 4 decimal places.

(Use 4 decimals behind the decimal point in your answer. Also, use a decimal point, *not* a comma and only round your *final* answer and not any intermediate results.)

Correct!

0.039

rrect Answers 0.039 (with margin: 0.0002)

Question 13

1 / 1 pts

For a sample of students the following data are available: whether the individual succeeded on the first attempt of the statistics exam, the logistics exam, and the mathematics exam.

You want to test whether these variables are related (i.e., dependent).

What is the correct approach for conducting your test? Give all correct answers.

- [ ] Paired 2-sample t-test
- [ ] Chi-squared goodness-of-fit test
- [ ] ANOVA
- [x] chi-squared-test — **Correct!**
- [ ] F-test
- [ ] None of the above

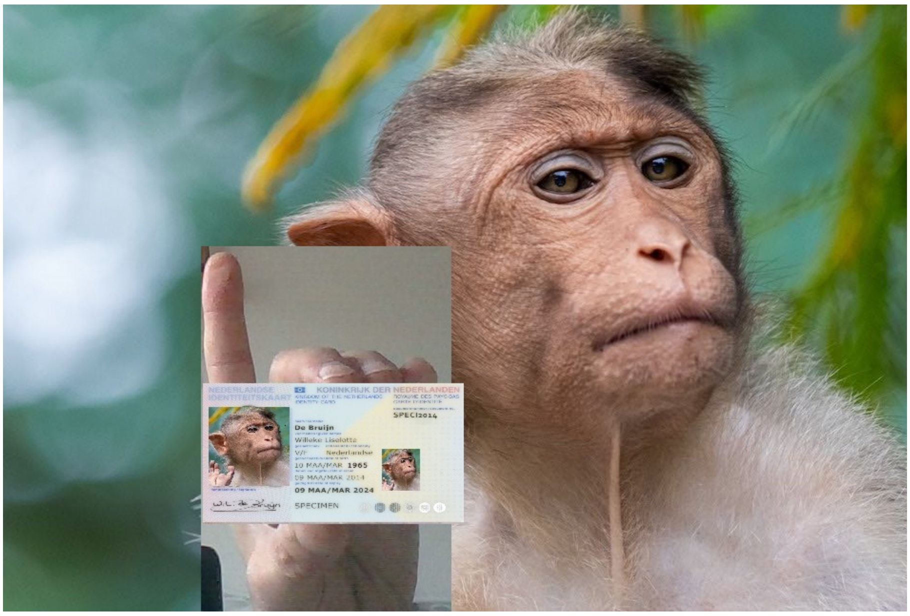
ıanswered

Question 14

Not yet graded / 1 pts

Upload a selfie such that:

- your face is clearly visible
- your ID with photo is clearly visible
- your hand holds the ID card index finger up, rest of fingers on the card, thumb at bottom.



Question 15

1 / 1 pts

In order to predict the outcome of the presidential election, a poll is held in the states of Florida (Fl) and New York (NY). In both states, a random sample of $n_{\text{Fl}} = n_{\text{NY}} = 60$ voters was asked if they would vote for the Democrat (D) or the Independent (I) candidate. In Florida, $x_{D,\text{Fl}} = 24$ voters preferred Democrat, against $x_{I,\text{Fl}} = 36$ who chose Independent. In New York however, $x_{D,\text{NY}} = 35$ preferred the Democrat, whereas $x_{I,\text{NY}} = 25$ had a preference for the Independent.

Define $\pi_{D,\text{Fl}}$ and $\pi_{D,\text{NY}}$ as the probabilities that a voter prefers the Democrat candidate over the Independent candidate in Florida, or in the second state, New York, respectively. These parameters are estimated by the fractions $p_{D,\text{Fl}}$ and $p_{D,\text{NY}}$ in the sample.

Likewise, population (β) and sample (b) regression parameters use the same indices.

Anne says that the sample sizes were too small to draw any conclusions. On the other hand, Piet is quite convinced by the difference. You are now asked to be the referee in this debate. Do these sample results show that indeed the Democrat party is less popular in Florida than in New York?

What is the H_0 hypothesis for this situation?

- $\beta_{D,\mathrm{Fl}} = \beta_{D,\mathrm{NY}}$
- $\beta_{D,\mathrm{Fl}} < \beta_{D,\mathrm{NY}}$
- $b_{D,\mathrm{Fl}} \neq b_{D,\mathrm{NY}}$
- $p_{D,\mathrm{Fl}} < p_{D,\mathrm{NY}}$
- $\pi_{D,\mathrm{Fl}} = \pi_{D,\mathrm{NY}}$
- **Correct!** $\pi_{D,\mathrm{Fl}} \geq \pi_{D,\mathrm{NY}}$
- None of the above

Question 16

1 / 1 pts

In order to predict the outcome of the presidential election, a poll is held in the states of Washington (W) and Florida (Fl). In both states, a random sample of $n_W = n_{\mathrm{Fl}} = 60$ voters was asked if they would vote for the Independent (I) or the Republican (R) candidate. In Washington, $x_{I,W} = 34$ voters preferred Independent, against $x_{R,W} = 26$ who chose Republican. In Florida however, $x_{I,\mathrm{Fl}} = 25$ preferred the Independent, whereas $x_{R,\mathrm{Fl}} = 35$ had a preference for the Republican.

Define $\pi_{I,W}$ and $\pi_{I,\mathrm{Fl}}$ as the probabilities that a voter prefers the Independent candidate over the Republican candidate in Washington, or in the second state, Florida, respectively. These parameters are estimated by the fractions $p_{I,W}$ and $p_{I,\text{Fl}}$ in the sample.

Anne says that the sample sizes were too small to draw any conclusions. On the other hand, Joe is quite convinced by the difference. You are now asked to be the referee in this debate. Do these sample results show that indeed the Independent party is more popular in Washington than in Florida?

If you want to test H_0: $\pi_{I,W} \leq \pi_{I,\text{Fl}}$ against H_1: $\pi_{I,W} > \pi_{I,\text{Fl}}$, what would be your test statistic and rejection region?

- Test statistic: $z = \frac{p - \pi_{I,W}}{\sqrt{\pi_{I,W}(1-\pi_{I,W})/n_W}}$, reject for small values

Correct!

- Test statistic: $z = \frac{p_{I,W} - p_{I,\text{Fl}} - (\pi_{I,W} - \pi_{I,\text{Fl}})}{\sqrt{p_{I,W}(1-p_{I,W})/n_W + p_{I,\text{Fl}}(1-p_{I,\text{Fl}})/n_{\text{Fl}}}}$, reject for large values
- Test statistic: $z = \frac{p - \pi_{I,\text{Fl}}}{\sqrt{\pi_{I,\text{Fl}}(1-\pi_{I,\text{Fl}})/n_{\text{Fl}}}}$, reject for small values
- Test statistic: $z = \frac{p_{I,W} - p_{I,\text{Fl}} - (\pi_{I,W} - \pi_{I,\text{Fl}})}{\sqrt{p_{I,W}(1-p_{I,W})/n_W + p_{I,\text{Fl}}(1-p_{I,\text{Fl}})/n_{\text{Fl}}}}$, reject for small values
- Test statistic: $z = \frac{p - \pi_{I,W}}{\sqrt{\pi_{I,W}(1-\pi_{I,W})/n_W}}$, reject for large values
- Test statistic: $z = \frac{p - \pi_{I,\text{Fl}}}{\sqrt{\pi_{I,\text{Fl}}(1-\pi_{I,\text{Fl}})/n_{\text{Fl}}}}$, reject for large values
- None of the above

Question 17

1 / 1 pts

In order to predict the efficacy of crop planting, a poll is held among the owners of farms in Noord Brabant (NB) and Friesland (Fr). In both provinces, a random sample of $n_{\mathrm{NB}} = n_{\mathrm{Fr}} = 50$ individuals were asked if they planned to increase their use of manure (M), or preferred to use more fertilizer (F). In Noord Brabant, $x_{M,\mathrm{NB}} = 29$ preferred manure, against $x_{F,\mathrm{NB}} = 21$ who chose fertilizer. In Friesland however, $x_{M,\mathrm{Fr}} = 23$ preferred the manure, whereas $x_{F,\mathrm{Fr}} = 27$ had a preference of using the fertilizer.

Define $\pi_{M,\mathrm{NB}}$ and $\pi_{M,\mathrm{Fr}}$ as the probabilities that a farmer prefers manure over the use of fertilizer in Noord Brabant, or in the second province, Friesland, respectively. These parameters are estimated by the fractions $p_{M,\mathrm{NB}}$ and $p_{M,\mathrm{Fr}}$ in the sample.

Anne says that the sample sizes were too small to draw any conclusions. On the other hand, Piet is quite convinced by the difference. You are now asked to be the referee in this debate. Do these sample results show that indeed manure is more popular in Noord Brabant than in Friesland?

We test H_0: $\pi_{M,\mathrm{NB}} \leq \pi_{M,\mathrm{Fr}}$, using test statistic $\frac{p_{M,\mathrm{NB}} - p_{M,\mathrm{Fr}} - (\pi_{M,\mathrm{NB}} - \pi_{M,\mathrm{Fr}})}{\sqrt{p_{M,\mathrm{NB}}(1-p_{M,\mathrm{NB}})/n_{\mathrm{NB}} + p_{M,\mathrm{Fr}}(1-p_{M,\mathrm{Fr}})/n_{\mathrm{Fr}}}}$, and reject for large values

If the null hypothesis is true, and the necessary assumptions fulfilled, the test statistic will be distributed as:

- t_{49}
- **Correct!** N(0,1)
- χ^2_{50}

- χ^2_{49}
- t_{100}
- t_{50}
- None of the above

Question 18

1 / 1 pts

In order to predict the outcome of the presidential election, a poll is held in the states of Texas (T) and Washington (W). In both states, a random sample of $n_T = n_W = 40$ voters was asked if they would vote for the Republican (R) or the Democrat (D) candidate. In Texas, $x_{R,T} = 16$ voters preferred Republican, against $x_{D,T} = 24$ who chose Democrat. In Washington however, $x_{R,W} = 21$ preferred the Republican, whereas $x_{D,W} = 19$ had a preference for the Democrat.

Define $\pi_{R,T}$ and $\pi_{R,W}$ as the probabilities that a voter prefers the Republican candidate over the Democrat candidate in Texas, or in the second state, Washington, respectively. These parameters are estimated by the fractions $p_{R,T}$ and $p_{R,W}$ in the sample.

Stijn says that the sample sizes were too small to draw any conclusions. On the other hand, Piet is quite convinced by the difference. You are now asked to be the referee in this debate. Do these sample results show that indeed the Republican party is less popular in Texas than in Washington?

We test H_0: $\pi_{R,T} \geq \pi_{R,W}$, using test statistic $z = \frac{p_{R,T} - p_{R,W} - (\pi_{R,T} - \pi_{R,W})}{\sqrt{p_{R,T}(1-p_{R,T})/n_T + p_{R,W}(1-p_{R,W})/n_W}}$, and reject for small values

If the null hypothesis is true, the test statistic is distributed as a N(0,1), if the following assumptions hold:

(select *ALL* options that apply; in the notation, an i stands for either R or D, and j for T or W, so i, j stands for all possible combinations of these)

Correct!

- [x] All $x_{i,j} \geq 10$
- [ ] All $\pi_{i,j} \geq 5$
- [ ] All $n_j \geq 5$
- [ ] All n are normally distributed
- [ ] All x are normally distributed
- [ ] All p are normally distributed
- [ ] None of the above

Question 19

1 / 1 pts

In order to predict the efficacy of crop planting, a poll is held among the owners of farms in North Holland (NH) and Friesland (Fr). In both provinces, a random sample of $n_{\mathrm{NH}} = n_{\mathrm{Fr}} = 55$ individuals were asked if they planned to increase their use of fertilizer (F), or preferred to use more manure (M). In North Holland, $x_{F,\mathrm{NH}} = 27$ preferred fertilizer, against $x_{M,\mathrm{NH}} = 28$ who chose manure. In Friesland however, $x_{F,\mathrm{Fr}} = 32$ preferred the fertilizer, whereas $x_{M,\mathrm{Fr}} = 23$ had a preference of using the manure.

Define $\pi_{F,\mathrm{NH}}$ and $\pi_{F,\mathrm{Fr}}$ as the probabilities that a farmer prefers fertilizer over the use of manure in North Holland, or in the second province, Friesland, respectively. These parameters are estimated by the fractions $p_{F,\mathrm{NH}}$ and $p_{F,\mathrm{Fr}}$ in the sample.

Wouter says that the sample sizes were too small to draw any conclusions. On the other hand, Piet is quite convinced by the difference. You are now asked to be the referee in this debate. Do these sample results show that indeed fertilizer is less popular in North Holland than in Friesland?

We test H_0: $\pi_{F,\mathrm{NH}} \geq \pi_{F,\mathrm{Fr}}$, using test statistic $z = \frac{p_{F,\mathrm{NH}} - p_{F,\mathrm{Fr}} - (\pi_{F,\mathrm{NH}} - \pi_{F,\mathrm{Fr}})}{\sqrt{p_{F,\mathrm{NH}}(1-p_{F,\mathrm{NH}})/n_{\mathrm{NH}} + p_{F,\mathrm{Fr}}(1-p_{F,\mathrm{Fr}})/n_{\mathrm{Fr}}}}$.

Calculate the test statistic z_{calc}, up to a precision of at least 3 digits after the decimal point.

Correct! -0.96

rrect Answers -0.959991 (with margin: 0.000991)

Question 20

1 / 1 pts

In order to predict the outcome of the presidential election, a poll is held in the states of California (C) and Washington (W). In both states, a random sample of $n_C = n_W = 40$ voters was asked if they would vote for the Independent (I) or the Democrat (D) candidate. In California, $x_{I,C} = 24$ voters preferred Independent, against $x_{D,C} = 16$ who chose Democrat. In Washington however, $x_{I,W} = 17$ preferred the Independent, whereas $x_{D,W} = 23$ had a preference for the Democrat.

Define $\pi_{I,C}$ and $\pi_{I,W}$ as the probabilities that a voter prefers the Independent candidate over the Democrat candidate in California, or in the second state, Washington, respectively. These parameters are estimated by the fractions $p_{I,C}$ and $p_{I,W}$ in the sample.

Jim says that the sample sizes were too small to draw any conclusions. On the other hand, Joe is quite convinced by the difference. You are now asked to be the referee in this debate. Do these sample results show that indeed the Independent party is more popular in California than in Washington?

We test H_0: $\pi_{I,C} \leq \pi_{I,W}$, using test statistic $z = \frac{p_{I,C} - p_{I,W} - (\pi_{I,C} - \pi_{I,W})}{\sqrt{p_{I,C}(1-p_{I,C})/n_C + p_{I,W}(1-p_{I,W})/n_W}}$, and with $\alpha = 0.01$.

Jim and Joe found that $z_{\text{calc}} = 1.590$.

What is the critical value of the test? Use a precision of at least 2 digits after the decimal point.

Correct! 2.3264

rrect Answers 2.326348 (with margin: 0.013652)

Question 21

1 / 1 pts

In order to predict the outcome of the presidential election, a poll is held in the states of Montana (M) and New York (NY). In both states, a random sample of $n_M = n_{\text{NY}} = 55$ voters was asked if they would vote for the Democrat (D) or the Republican (R) candidate. In Montana, $x_{D,M} = 23$ voters preferred Democrat, against $x_{R,M} = 32$ who chose Republican. In New York however, $x_{D,\text{NY}} = 28$ preferred the Democrat, whereas $x_{R,\text{NY}} = 27$ had a preference for the Republican.

Define $\pi_{D,M}$ and $\pi_{D,\text{NY}}$ as the probabilities that a voter prefers the Democrat candidate over the Republican candidate in Montana, or in the second state, New York, respectively. These parameters are estimated by the fractions $p_{D,M}$ and $p_{D,\text{NY}}$ in the sample.

Joe says that the sample sizes were too small to draw any conclusions. On the other hand, Stijn is quite convinced by the difference. You are now asked to be the referee in this debate. Do these sample results show that indeed the Democrat party is less popular in Montana than in New York?

We test H_0: $\pi_{D,M} \geq \pi_{D,\text{NY}}$, using test statistic $z = \frac{p_{D,M} - p_{D,\text{NY}} - (\pi_{D,M} - \pi_{D,\text{NY}})}{\sqrt{p_{D,M}(1-p_{D,M})/n_M + p_{D,\text{NY}}(1-p_{D,\text{NY}})/n_{\text{NY}}}}$, and with $\alpha = 0.1$.

Joe and Stijn found that $z_{\text{calc}} = -0.960$, and $z_{\text{crit}} = -1.282$.

What is the conclusion they should draw?

- Do not reject H_0, as p-value $\leq \alpha$

Correct!

- Do not reject H_0, as $z_{\text{calc}} > z_{\text{crit}}$
- Reject H_1, as $z_{\text{calc}} < z_{\text{crit}}$
- Reject H_0, as p-value $> \alpha$
- Reject H_0, as p-value $< \alpha$
- Reject H_1, as $z_{\text{calc}} \leq z_{\text{crit}}$
- Reject H_1, as p-value $\leq \alpha$
- Do not reject H_1, as $z_{\text{calc}} \geq z_{\text{crit}}$
- None of the above

Question 22

1 / 1 pts

In order to predict the outcome of the presidential election, a poll is held in the states of Florida (Fl) and Texas (T). In both states, a random sample of $n_{\text{Fl}} = n_T = 50$ voters was asked if they would vote for the Republican (R) or the Independent (I) candidate. In Florida, $x_{R,\text{Fl}} = 28$ voters preferred Republican, against $x_{I,\text{Fl}} = 22$ who chose Independent. In Texas however, $x_{R,T} = 21$ preferred the Republican, whereas $x_{I,T} = 29$ had a preference for the Independent.

Floor says that the sample sizes were too small to draw any conclusions. On the other hand, Jim is quite convinced by the difference. You are now asked to be the referee in this debate. Do these sample results indeed show a difference in popularity of the Republican party between Florida and Texas?

Instead of testing $H_0 : \pi_{R,\text{Fl}} = \pi_{R,T}$, using the test statistic $z = \frac{p_{R,\text{Fl}} - p_{R,T} - (\pi_{R,\text{Fl}} - \pi_{R,T})}{\sqrt{p_{R,\text{Fl}}(1-p_{R,\text{Fl}})/n_{\text{Fl}} + p_{R,T}(1-p_{R,T})/n_T}}$, Floor and Jim now would like to see a two-sided confidence interval for the difference $d = \pi_{R,\text{Fl}} - \pi_{R,T}$.

For this purpose, they calculated the variance used in the test statistic $\sigma_d^2 = p_{R,\text{Fl}}(1 - p_{R,\text{Fl}})/n_{\text{Fl}} + p_{R,T}(1 - p_{R,T})/n_T = 0.009800$

Give the upper bound of the 90% **two-sided** confidence interval, up to 4 digits after the decimal point.

Correct! 0.3028

rrect Answers 0.302832 (with margin: 6.8e-05)

Question 23

0 / 1 pts

In the following questions, the numbers and setting may change, though questions are related. Use the numbers of the *question on screen* to answer the question.

In California, we investigate the lot prices. These lots (bouwkavels) are empty pieces of land where a new house may be build. For this purpose, we have taken a random sample of size 38 of prices of lots. The price may be linked to the area of lot, and its location, in either a rural zone, suburban or a city. Both Price and Area are measured as continuous variables, whereas Rural and City are dummy variables related to the location. The data were processed in R, with output for a restricted and an unrestricted regression model in Tables 1 and 2.

```
Table 1: Restricted model
              Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  -133.678     167.750   -0.797     0.431
    df$Area     1.937       0.859    2.254     0.030

Residual standard error: 139.2 on 36 degrees of freedom
Multiple R-squared: 0.1237,     Adjusted R-squared: 0.09932
F-statistic: 5.080 on 1 and 36 DF,  p-value: 0.03039
```

```
Table 2: Unrestricted model
              Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)   -15.479      32.833   -0.471     0.640
    df$Area     1.103       0.162    6.804     0.000
   df$Rural  -103.200      10.455   -9.871     0.000
    df$City   221.268      10.028   22.064     0.000

Residual standard error: 25.6 on 34 degrees of freedom
Multiple R-squared: 0.9721,     Adjusted R-squared: 0.9696
F-statistic: 394.647 on 3 and 34 DF,  p-value: < 2.22e-16
```

What is the theoretical model corresponding to the output for the restricted model?

- $\hat{y}=167.750+0.859x_1+\hat{e}$
- $\hat{y}=-15.479+1.103x_1+-103.200x_2+221.268x_3$
- rrect answer — $y=\beta_0+\beta_1 x_1+\varepsilon$
- $\hat{y}=167.750+0.859x_1$

$\hat{y} = \beta_0 + \beta_1 x_1$

$y = \beta_0 + \varepsilon$

ı Answered

None of the above

Question 24

1 / 1 pts

In the following questions, the numbers and setting may change, though questions are related. Use the numbers of the *question on screen* to answer the question.

In a number of counties in the US, we investigate the corona infections. For this purpose, we have taken a random sample of size 37 of number of inhabitants infected by the corona virus. The number of infected people may be linked to the population in the county (in '000s of people), and the location of the county, in either a rural zone, suburban or a city. Both Infected and Inhabitants are measured as continuous variables, whereas Rural and City are dummy variables related to the location. The data were processed in R, with output for a restricted and an unrestricted regression model in Tables 1 and 2.

```
Table 1: Restricted model
                Estimate  Std. Error  t value  Pr(>|t|)
   (Intercept)   329.878     142.227    2.319     0.026
df$Inhabitants    -0.456       0.728   -0.626     0.535

Residual standard error: 129.3 on 35 degrees of freedom
Multiple R-squared: 0.01109,    Adjusted R-squared: -0.0171
7
F-statistic: 0.392 on 1 and 35 DF,  p-value: 0.5351
```

```
Table 2: Unrestricted model
                Estimate  Std. Error  t value  Pr(>|t|)
   (Intercept)    48.397      26.457    1.829     0.076
```

```
df$Inhabitants     0.850       0.132    6.449     0.000
      df$Rural  -113.708       9.008  -12.623     0.000
       df$City   203.927       9.133   22.328     0.000

Residual standard error: 22.4 on 33 degrees of freedom
Multiple R-squared: 0.9721,     Adjusted R-squared: 0.9696
F-statistic: 383.204 on 3 and 33 DF,  p-value: < 2.22e-16
```

What descriptions of variables belong to the model description for the unrestricted model? Select *all* that apply.

- [ ] β_0: expected increase in Number of infected patients

Correct!
- [x] x_1: Number of inhabitants (x 1000)

Correct!
- [x] x_3: City location, dummy

Correct!
- [x] x_2: Rural location, dummy

- [ ] β_3: expected increase in Number of infected patients, when City location, dummy increases by 1 unit

- [ ] β_3: coefficient of City location, dummy

- [ ] β_4: expected increase in Number of infected patients, when Suburban location, dummy increases by 1 unit

Correct!
- [x] y: Number of infected patients

Question 25

1 / 1 pts

In the following questions, the numbers and setting may change, though questions are related. Use the numbers of the *question on screen* to answer the question.

In a number of counties in the US, we investigate the corona infections. For this purpose, we have taken a random sample of size 41 of number of inhabitants infected by the corona virus. The number of infected people may be linked to the population in the county (in '000s of people), and the location of the county, in either a rural zone, suburban or a city. Both Infected and Inhabitants are measured as continuous variables, whereas Rural and City are dummy variables related to the location. The data were processed in R, with output for a restricted and an unrestricted regression model in Tables 1 and 2.

```
Table 1: Restricted model
                 Estimate  Std. Error  t value  Pr(>|t|)
   (Intercept)    85.872     145.747    0.589     0.559
df$Inhabitants     0.828       0.716    1.156     0.255

Residual standard error: 141.8 on 39 degrees of freedom
Multiple R-squared: 0.03315,    Adjusted R-squared: 0.00836
4
F-statistic: 1.337 on 1 and 39 DF,  p-value: 0.2545
```

```
Table 2: Unrestricted model
                 Estimate  Std. Error  t value  Pr(>|t|)
   (Intercept)    12.071      32.878    0.367     0.716
df$Inhabitants     1.011       0.153    6.609     0.000
      df$Rural  -113.594      11.611   -9.783     0.000
       df$City   214.046      11.412   18.756     0.000

Residual standard error: 29.8 on 37 degrees of freedom
Multiple R-squared: 0.9594,     Adjusted R-squared: 0.9561
F-statistic: 291.607 on 3 and 37 DF,  p-value: < 2.22e-16
```

Use the output in the restricted model in order to predict the value of 'Infected' given that 'Inhabitants'=90 when the location is Suburban

Give your answer with a precision of 3 decimals after the decimal point.

Correct!

160.392

rrect Answers 160.392 (with margin: 0.001)

Question 26

1 / 1 pts

In the following questions, the numbers and setting may change, though questions are related. Use the numbers of the *question on screen* to answer the question.

In a number of counties in the US, we investigate the corona infections. For this purpose, we have taken a random sample of size 31 of number of inhabitants infected by the corona virus. The number of infected people may be linked to the population in the county (in '000s of people), and the location of the county, in either a rural zone, suburban or a city. Both Infected and Inhabitants are measured as continuous variables, whereas Rural and City are dummy variables related to the location. The data were processed in R, with output for a restricted and an unrestricted regression model in Tables 1 and 2.

```
Table 1: Restricted model
                Estimate  Std. Error  t value  Pr(>|t|)
   (Intercept)  -149.514     202.097   -0.740     0.465
df$Inhabitants     2.010       1.014    1.983     0.057

Residual standard error: 145.2 on 29 degrees of freedom
Multiple R-squared: 0.1194,     Adjusted R-squared: 0.08902
F-statistic: 3.932 on 1 and 29 DF,  p-value: 0.05693
```

```
Table 2: Unrestricted model
                Estimate  Std. Error  t value  Pr(>|t|)
   (Intercept)    59.323      32.478    1.827     0.079
df$Inhabitants     0.782       0.163    4.781     0.000
      df$Rural  -121.646       9.954  -12.221     0.000
       df$City   227.149      10.146   22.388     0.000

Residual standard error: 22.8 on 27 degrees of freedom
Multiple R-squared: 0.9798,     Adjusted R-squared: 0.9776
F-statistic: 436.882 on 3 and 27 DF,  p-value: < 2.22e-16
```

John claims that every extra 1000 inhabitants will increase the corona infections with more than 1.8. Can John's claim be proven by the results of the restricted model?

What is the relevant alternative hypothesis?

Correct!

- (selected) $H_1 : \beta_1 > 1.8$
- $H_1 : \beta_1 \neq 1.8$
- $H_1 : \beta_0 < 1.8$
- $H_0 : \beta_1 = 1.8$
- $H_0 : \beta_0 = 1.8$
- $H_0 : \beta_0 \geq 1.8$
- $H_0 : \beta_1 \geq 1.8$
- $H_0 : \beta_0 \leq 1.8$
- None of the above

Question 27

1 / 1 pts

In the following questions, the numbers and setting may change, though questions are related. Use the numbers of the *question on screen* to answer the question.

In a number of counties in the US, we investigate the corona infections. For this purpose, we have taken a random sample of size 40 of number of inhabitants infected by the corona virus. The number of infected people may be linked to the population in the county (in '000s of people), and the location of the county, in either a rural zone, suburban or a city. Both Infected and Inhabitants are measured as continuous variables, whereas Rural and City are dummy variables related to the location. The data were processed in R, with output for a restricted and an unrestricted regression model in Tables 1 and 2.

```
Table 1: Restricted model
                Estimate  Std. Error  t value  Pr(>|t|)
   (Intercept)  -228.301     186.591   -1.224     0.229
df$Inhabitants     2.360       0.921    2.563     0.014

Residual standard error: 131.9 on 38 degrees of freedom
Multiple R-squared: 0.1474,     Adjusted R-squared: 0.125
F-statistic: 6.571 on 1 and 38 DF,  p-value: 0.01444
```

```
Table 2: Unrestricted model
                Estimate  Std. Error  t value  Pr(>|t|)
   (Intercept)   -33.335      36.228   -0.920     0.364
df$Inhabitants     1.213       0.173    6.989     0.000
      df$Rural  -101.235       9.490  -10.668     0.000
       df$City   212.426       9.222   23.034     0.000

Residual standard error: 23.9 on 36 degrees of freedom
Multiple R-squared: 0.9734,     Adjusted R-squared: 0.9712
F-statistic: 438.937 on 3 and 36 DF,  p-value: < 2.22e-16
```

John claims that every extra 1000 inhabitants will increase the corona infections with something different from 1.2 in the unrestricted model. To try and prove his point, he wants to test the null hypothesis $H_0 : \beta_1 = 1.2$ against the alternative $H_1 : \beta_1 \neq 1.2$. What is the corresponding test statistic?

Select *all* possible answers

- [ ] $t = \beta_1 / s_{\beta_1}$
- [ ] β_1
- [ ] $t = b_0 / s_{b_0}$
- [ ] $t = b_1 / s_{b_1}^2$

- [ ] $t = b_0 / s_{b_0}^2$
- [ ] b_0

Correct!
- [x] $t = (b_1 - 1.2) / s_{b_1}$
- [ ] $t = (\beta_1 - 1.2) / s_{\beta_1}$
- [ ] None of the above

Question 28

1 / 1 pts

In the following questions, the numbers and setting may change, though questions are related. Use the numbers of the *question on screen* to answer the question.

In a number of counties in the US, we investigate the corona infections. For this purpose, we have taken a random sample of size 37 of number of inhabitants infected by the corona virus. The number of infected people may be linked to the population in the county (in '000s of people), and the location of the county, in either a rural zone, suburban or a city. Both Infected and Inhabitants are measured as continuous variables, whereas Rural and City are dummy variables related to the location. The data were processed in R, with output for a restricted and an unrestricted regression model in Tables 1 and 2.

```
Table 1: Restricted model
                 Estimate  Std. Error  t value  Pr(>|t|)
   (Intercept)   -99.334     144.281   -0.688     0.496
df$Inhabitants     1.716       0.715    2.400     0.022

Residual standard error: 140.9 on 35 degrees of freedom
Multiple R-squared: 0.1413,     Adjusted R-squared: 0.1167
F-statistic: 5.758 on 1 and 35 DF,  p-value: 0.02187
```

Table 2: Unrestricted model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-72.892	27.940	-2.609	0.014
df$Inhabitants	1.432	0.130	11.017	0.000
df$Rural	-119.454	10.252	-11.651	0.000
df$City	211.980	10.060	21.071	0.000

Residual standard error: 24.9 on 33 degrees of freedom
Multiple R-squared: 0.9746, Adjusted R-squared: 0.9723
F-statistic: 422.872 on 3 and 33 DF, p-value: < 2.22e-16

John wants to test the null hypothesis $H_0 : \beta_1 \geq 1.5$ in the restricted model, with test statistic $t = (b_1 - 1.5)/s_{b_1}$.

In this case, the test should reject:

- [ ] in the middle
- [ ] never
- [ ] on the left and right side
- [ ] cannot tell given the information
- [x] None of the above

Correct!

Question 29

1 / 1 pts

In the following questions, the numbers and setting may change, though questions are related. Use the numbers of the *question on screen* to answer the question.

In a number of US states, we investigate the supermarket sales. For this purpose, we have taken a random sample of size 41 of sales of supermarkets. The sales volume may be linked to the area of the parking lot, and the location of the supermarket, in either a rural zone, suburban or a city. Both

Sales and Area are measured as continuous variables, whereas Rural and City are dummy variables related to the location. The data were processed in R, with output for a restricted and an unrestricted regression model in Tables 1 and 2.

```
Table 1: Restricted model
              Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)    77.380     133.251    0.581     0.565
    df$Area     0.854       0.690    1.236     0.224

Residual standard error: 141.8 on 39 degrees of freedom
Multiple R-squared: 0.03772,    Adjusted R-squared: 0.01305
F-statistic: 1.529 on 1 and 39 DF,  p-value: 0.2237
```

```
Table 2: Unrestricted model
              Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)    53.179      22.921    2.320     0.026
    df$Area     0.852       0.116    7.349     0.000
   df$Rural  -135.834       9.173  -14.808     0.000
    df$City   197.803       9.002   21.974     0.000

Residual standard error: 23.8 on 37 degrees of freedom
Multiple R-squared: 0.9742,     Adjusted R-squared: 0.9722
F-statistic: 466.542 on 3 and 37 DF,  p-value: < 2.22e-16
```

John wants to test the null hypothesis $H_0 : \beta_1 \leq 0.8$ in the restricted model, with test statistic $t = (b_1 - 0.8)/s_{b_1}$.

In this case, if the null hypothesis is true, the test statistic t is distributed as:

- [ ] an F(1,39) distribution
- [ ] a t_{37} distribution
- [ ] a χ^2_{37} distribution
- [ ] a t_{38} distribution
- [ ] a χ^2_{38} distribution

- [ ] a χ^2_{39} distribution
- [ ] a t_{40} distribution
- [ ] a t_{41} distribution

Correct!

- [x] None of the above

Question 30

1 / 1 pts

In the following questions, the numbers and setting may change, though questions are related. Use the numbers of the *question on screen* to answer the question.

In a number of counties in the US, we investigate the corona infections. For this purpose, we have taken a random sample of size 32 of number of inhabitants infected by the corona virus. The number of infected people may be linked to the population in the county (in '000s of people), and the location of the county, in either a rural zone, suburban or a city. Both Infected and Inhabitants are measured as continuous variables, whereas Rural and City are dummy variables related to the location. The data were processed in R, with output for a restricted and an unrestricted regression model in Tables 1 and 2.

```
Table 1: Restricted model
                Estimate  Std. Error  t value  Pr(>|t|)
   (Intercept)   262.919     167.761    1.567     0.128
df$Inhabitants    -0.017       0.810   -0.021     0.983

Residual standard error: 125.7 on 30 degrees of freedom
Multiple R-squared: 1.473e-05,  Adjusted R-squared: -0.0333
2
F-statistic: 0.000 on 1 and 30 DF,  p-value: 0.9834
```

```
Table 2: Unrestricted model
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	67.782	51.057	1.328	0.195
df$Inhabitants	0.759	0.228	3.324	0.002
df$Rural	-96.745	14.869	-6.507	0.000
df$City	192.149	15.046	12.771	0.000

```
Residual standard error: 32.7 on 28 degrees of freedom
Multiple R-squared: 0.9368,     Adjusted R-squared: 0.93
F-statistic: 138.314 on 3 and 28 DF,  p-value: < 2.22e-16
```

John tested the null hypothesis $H_0 : \beta_1 = 0.1$ in the restricted model, and finds $t = (b_1 - 0.1)/s_{b_1} = -0.144444$

What is the corresponding critical value, at level $\alpha = 0.1$?

(Use 3 decimals behind the decimal point in your answer. Also, use a decimal point, *not* a comma and only round your *final* answer and not any intermediate results.)

(If there are two critical values, report the highest one)

Correct! 1.6973

rrect Answers 1.697261 (with margin: 0.000739)

Question 31

1 / 1 pts

In the following questions, the numbers and setting may change, though questions are related. Use the numbers of the *question on screen* to answer the question.

In California, we investigate the lot prices. These lots (bouwkavels) are empty pieces of land where a new house may be build. For this purpose, we have taken a random sample of size 37 of prices of lots. The price may be linked to the area of lot, and its location, in either a rural zone, suburban or a city. Both Price and Area are measured as continuous variables, whereas Rural and City are dummy variables related to the location. The data were processed in R, with output for a restricted and an unrestricted regression model in Tables 1 and 2.

```
Table 1: Restricted model
              Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  -124.234     129.108   -0.962     0.343
    df$Area     1.803       0.634    2.842     0.007

Residual standard error: 133.4 on 35 degrees of freedom
Multiple R-squared: 0.1875,     Adjusted R-squared: 0.1643
F-statistic: 8.079 on 1 and 35 DF,  p-value: 0.007423
```

```
Table 2: Unrestricted model
              Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)    -1.869      25.099   -0.074     0.941
    df$Area     1.009       0.124    8.122     0.000
   df$Rural  -101.538      10.176   -9.978     0.000
    df$City   215.297      10.420   20.661     0.000

Residual standard error: 25.4 on 33 degrees of freedom
Multiple R-squared: 0.9722,     Adjusted R-squared: 0.9697
F-statistic: 384.912 on 3 and 33 DF,  p-value: < 2.22e-16
```

John tested the null hypothesis $H_0{:}\beta_1 \geq 1.9$ in the restricted model, and finds $t = (b_1 - 1.9)/s_{b_1} = -0.152997$.

As critical value, he finds $t_{\text{crit}} = -1.306$.

What is his conclusion?

- reject
- accept
- do not reject — Correct!
- None of the above

Question 32

1 / 1 pts

Before submitting your exam, choose one of the statements below that applies to you.

- I used the help of people from the internet during this exam.

Correct!

- I made this exam by myself without any outside help for the entire duration of the exam.
- I used the help of some friends during this exam.

Quiz score: **28** out of 32