Bob's telepathic gifts

Show block intro

Bob pretends to have telepathic gifts in the sense that if a card is drawn randomly from a set with as many red as black cards, he is able to name the correct color without looking at it. We proceed as follows: we let him guess 100 consecutive times, where the drawn card is put back every time. There were 50 red cards and 50 black cards drawn from a shuffled deck, Bob guessed the color of 30 black cards and 35 red cards correctly.

When answering the below questions, you can use the quantiles *qnorm(0.95)=1.64* and *qnorm(0.975)=1.96.*

Your answer is partially correct (See correct answer below)

- The estimate of the probability p that Bob names the right color is [0.65] (give a numerical value rounded to two decimals), 95% confidence interval for p is [0.56, 0.74] (choose from the drop down menu).
- Bob claims that he has probability at least 0.6 of naming the correct color. To verify his claim, we perform an appropriate test. Choose the correct claim from the drop down menu: [The proportion test yields the statistic value T=1.02 < 1.96=qnorm(0.975) so that the claim is not confirmed.].
- The minimal sample size needed to provide that the length of the 95%-confidence interval for p is at most 0.2 is [88] (give an integer number).

Bob's telepathic gifts

Show block intro

Bob pretends to have telepathic gifts in the sense that if a card is drawn randomly from a set with as many red as black cards, he is able to name the correct color without looking at it. We proceed as follows: we let him guess 100 consecutive times, where the drawn card is put back every time. There were 50 red cards and 50 black cards drawn from a shuffled deck, Bob guessed the color of 30 black cards and 35 red cards correctly.

A friend of Bob, Alice, claims that whether Bob guesses the color or not depends on the card color. To test the claim, represent the data in the form of contingency table

	guessed	not guessed
black	V1	V2
red	V2	V4

Your answer is partially correct **(See correct answer below)**

Determine the values of this contingency table V1= 30 , V2= 20 , V3= 35 , V4= 15 . Next, suppose we put those values in the matrix *m=matrix(c(V1,V2,V3,V4),byrow=TRUE,ncol=2,nrow=2)* and want to apply a suitable chi-square test with significance level $\alpha = 0.05$. You may use the quantiles: *qchisq(0.95,1)=3.84*, *qchisq(0.975,1)=5.02*, *qchisq(0.95,2)=5.99*, *qchisq(0.05,1)=0.004*. We compute the value of the chi-squared test statistics X^2=0.7033 and conclude that the card color *does not affect* the Bob's ability to guess the color.

Choose the best test to verify the claim of Alice from the following drop down menu: the Fisher test .

0.31 / 1 pt.

Bob's telepathic gifts Show block intro

Bob pretends to have telepathic gifts in the sense that if a card is drawn randomly from a set with as many red as black cards, he is able to name the correct color without looking at it. We proceed as follows: we let him guess 100 consecutive times, where the drawn card is put back every time.

We have collected our data in the data frame *bob.txt*, where column *guessed* indicates whether Bob guessed the color or not ("yes" or "no"), *time* is the time (in seconds) that Bob spent when guessing the color.

Suppose we want to study the influence of factor *guessed* on the variable *time* (that is, *time* becomes now the response variable with factor *guessed*). Choose the correct claim(s).

Your answer is partially correct (See correct answer below)

- [x] This problem can be addressed by the Kolmogorov-Smirnov test.
- [ ] The Friedman test is relevant for this problem.
- [ ] This problem can be addressed by the Shapiro-Wilk test.
- [x] This problem can be addressed by the one-way ANOVA model.

0.17 / 1 pt.

Bob's telepathic gifts Show block intro

Bob pretends to have telepathic gifts in the sense that if a card is drawn randomly from a set with as many red as black cards, he is able to name the correct color without looking at it. We proceed as follows: we let him guess 100 consecutive times, where the drawn card is put back every time. There were 50 red cards and 50 black cards drawn from a shuffled deck, Bob guessed the color of 30 black cards and 35 red cards correctly.

Suppose we have collected our data in the data frame *bob.txt,* where column *color* is the color of the drawn card ("red" or "black"), *guessed* indicates whether Bob guessed the color or not ("yes" or "no"), *time* is the time (in seconds) that Bob spent when guessing the color.

We run *bob1=glm(guessed~color+time,family=binomial,data=bob); drop1(bob1,test="Chisq")* to obtain the output

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		127.59	133.59		
color	1	128.68	132.68	1.08746	0.297
time	1	128.39	132.39	0.79379	0.373

We create the model with the intercept term only: *bob2=glm(guessed~1,data=bob,family=binomial).* Next we run *summary(bob2)* to obtain the following output

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.6190	0.2097	2.953	0.00315 **

Choose the correct claim(s) on the basis of the above outputs.

Your answer is partially correct (See correct answer below)

- [x] The both variables *color* and *time* have no effect on *guessed*.
- [ ] If we use the model *bob2* for predicting the probability of guessing the red color for *time*=2, we obtain 0.7.
- [x] If we use the model *bob2* for predicting the probability of guessing the black color for *time*=2, we obtain 0.65.

0.5 / 1 pt.

Bob's telepathic gifts Show block intro

Bob pretends to have telepathic gifts in the sense that if a card is drawn randomly from a set with as many red as black cards, he is able to name the correct color without looking at it. We proceed as follows: we let him guess 100 consecutive times, where the drawn card is put back every time.

Suppose we have collected our data in the data frame *bob.txt*, where column *color* is the color of the drawn card ("red" or "black"), *guessed* indicates whether Bob guessed the color or not ("yes" or "no"), *time* is the time (in seconds) that Bob spent when guessing the color. A friend of Bob, Alice, thinks that whether Bob guesses the color depends on the both variables *color* and *time*. She also believes that the variables *time* and *color* interact.

We first run R-commands *alice1=glm(guessed~time*color,data=bob,family=binomial); anova(alice1,test="Chisq")* and obtain the following output

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			99	129.49	
time	1	0.80806	98	128.68	0.3687
color	1	1.08746	97	127.59	0.2970
time:color	1	0.22697	96	127.37	0.6338

Then we run *alice2=glm(guessed~time+color,data=bob,family=binomial); anova(alice2,test="Chisq")* and obtain the following output

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			99	129.49	
time	1	0.80806	98	128.68	0.3687
color	1	1.08746	97	127.59	0.2970

On the basis of the above outputs, choose the correct claim(s).

Your answer is partially correct (See correct answer below)

- [x] There is no main effect of *color*.
- [ ] There is a main effect of *time*.
- [x] There is no interaction between factor *color* and variable *time*.
- [ ] Both *time* and *color* affect *guessed*.

0.43 / 1 pt.

Varia

Show block intro

Consider the models *mod1=lm(y~f1+x1)* and *mod2=lm(y~x1+f1)*, where *f1* is a factor and *x1* is a continuous variable. Choose the correct claim(s).

Your answer is partially correct (See correct answer below)

- [x] The p-values for *x1* in *anova(mod1)* and *summary(mod2)* are the same.
- [x] The p-values for *f1* in *anova(mod2)* and *drop1(mod1)* are the same.
- [ ] The p-values for *x1* in *anova(mod2)* and *summary(mod2)* are the same.
- [ ] The p-values for *x1* in *anova(mod1)* and *drop1(mod2)* may be different.
- [x] The p-values for *f1* in *anova(mod1)* and *drop1(mod1)* may be different.

0.13 / 1 pt.

Birthweights Show block intro

Suppose we observed the birth weights (variable *y* in grams) and lengths of pregnancy (variable *x* in weeks) of 32 babies and fitted a simple linear model with covariate *x* and response variable *y*. The R-output of the regression analysis *summary(lm(y~x,data=data))* is partially presented below.

```
Coefficients:
              Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)  -2037.00       V1       -4.089      V2
x              130.82    12.86         V3        V4
---
Residual standard error: 167.3 on 30 degrees of freedom
Multiple R-squared:  0.7752,    Adjusted R-squared:  0.7677
F-statistic: 103.4 on 1 and 30 DF,  p-value: 3.085e-11
```

What missing information in the partial R-output can we recover by using **only the above R-output**?

Your answer is partially correct (See correct answer below)

- ☐ We can recover the value V2.
- ☒ We can recover the value V3.
- ☒ We can recover the value V1.
- ☐ We can recover the value V4.

0.31 / 1 pt.

Birthweights Show block intro

Suppose we observed the birth weights and the smoking status of the mother ("yes","no") for 32 newborn babies and we want to test the claim that smoking (of the mother) leads to lower birth weight of the baby. Let *y1* be the sample of the birth weights of babies of smoking mothers and *y2* of non-smoking mothers.

Choose the correct claim(s).

Your answer is partially correct (See correct answer below)

- [x] The Mann-Whitney test as *wilcox.test(y1,y2,alt="l")* is relevant in this situation.
- [ ] The permutation tests for two paired samples is relevant in this situation.
- [x] The Kolmogorov-Smirnov test as *ks.test(y1,y2,alt="g")* is relevant in this situation.
- [ ] The sign test and the Wilcoxon signed rank test are relevant in this situation.

0 / 1 pt.

Varia Show block intro

The following failure and censoring times (in operating hours) were recorded on 12 turbine vanes: 142, 149, 329, 345+, 560, 805, 1130+, 1720, 2480+, 4210+, 5230, 6890 (+ indicating censored observation). Censoring was a result of failure mode other than wear-out. We are interested in the distribution of the lifetime T of a turbine vane (survival function S(t)=P(T>t)).

Determine the default 95% confidence interval for S(565). You can use the normal upper quantiles $z_{0.025}$=*qnorm(0.975)*=1.96 and $z_{0.05}$=*qnorm(0.95)*=1.64.

Your answer is incorrect (See correct answer below)

- (●) [0.432,0.997]
- ( ) [0.437,0.987]
- ( ) [0.427,0.987]
- ( ) impossible to determine
- ( ) [0.437,0.992]

0 / 1 pt.

Varia Show block intro

The following failure and censoring times (in operating hours) were recorded on 12 turbine vanes: 142, 149, 329, 345+, 560, 805, 1130+, 1720, 2480+, 4210+, 5230, 6890
(+ indicating censored observation). Censoring was a result of failure mode other than wear-out. We are interested in the distribution of the lifetime T of a turbine vane (survival function S(t)=P(T>t)).

Your answer is incorrect **(See correct answer below)**

- At time t=565 the value at risk is n(565)= [7] (give an integer number) and the number of failures at time t=565 is [0] (give an integer number).
- The Kaplan-Meier estimator of the survival function S at time t=565, is [0.656] (give a number rounded to 3 decimals).

0.31 / 1

Varia

Show block intro

Suppose we observe some counts treated as response variable and some explanatory variables, and we assume a Poisson regression model. Choose the correct claim(s).

Your answer is partially correct (See correct answer below)

- [x] The variances of the observations can be estimated.
- [ ] A normal QQ-plot of the residuals can be used for the model diagnostics.
- [x] The distributions of the observations are assumed to be Poisson.
- [ ] The observations are assumed to have the same distribution.