



Försättsblad Prov Original

Kurskod	ST001G	Provkod	T102	Tentamensdatum	2018 - 08 - 21
Kursnamn	Statistik GR (A), Grundkurs				
Provnamn	Skriftlig tentamen				
Ort	Östersund				
Termin	H18				
Ämne	Statistik				

ST001G

Mittuniversitetet
EJV, Avdelningen för ekonomivetenskaper och juridik

Tentamen statistik A1 15 hp

Dag: 2018-08-21
Tid till förfogande: 5 timmar

OBLIGATORISK LÄSNING

Hjälpmedel: Miniräknare (valfri)
Tabell- och formelsamling

Anvisningar: Redovisa tydligt tankegången i lösningarna. Visa **samtliga** uträkningar där inget annat är angivet.

Redovisa varje uppgift på separat ark.

Betygsgränser

A 45-50
B 40-44
C 35-39
D 30-34
E 25-29
F 0-24

Fx ges inte i kursen. Du kan alltså **inte** få komplettering upp till E. Detta är ett principbeslut inom ämnet företagsekonomi där statistik kursen ingår.

Uppgift 1

(8 p)

Borde SVT reklamfinansieras? Denna fråga ställdes till 500 slumpmässigt utvalda innehavare av TV-licens. Ett frågeformulär skickades via posten ("snigelposten"). Resultatet av studien sammanfattas i korstabellen som visar antal.

	Kvinna	Man
Ja	165	115
Nej	80	140

Beräkningarna ska redovisas och svaren tydligt framgå.

- Beräkna ett konfidensintervall (95 %) för andel **Ja** för populationen TV-licensinnehavare.
- Vad säger KI:s gränser i just detta sammanhang? Tolka utförligt!
- Diskutera förutsättningar för konfidensintervallet ovan.

Uppgift 2

(8 p)

Anta att vi vill testa om den genomsnittliga försäljningen för Kim är mindre än \$1500. Ett slumpmässigt stickprov om 30 försäljningar (under dem tre senaste månaderna som Kim gjort) visar att medelvärdet är \$1439 och standardavvikelsen är \$ 1330. Kan vi dra slutsatsen att Kim i genomsnitt säljer mindre än \$1500?

Beräkningarna ska redovisas och svaren tydligt framgå.

Svara på frågan genom att göra ett hypotestest. **Motivera** ditt svar och glöm inte att formulera hypoteser och diskutera förutsättningarna.

Uppgift 3

(8 p)

Fruktodlaren, Mwanajuma, i Zanzibar säljer 50 kg mango och 20 kg bananer. Marknadspriset för mango (per kg) varje dag är normalslumpvariabel med medelvärde 5 000 Tzs (Tanzanisk shilling) och standardavvikelsen 900 Tzs. Motsvarande för bananer är medelvärde 3 000 Tzs och standardavvikelsen 400 Tzs. Det kostar henne 16 000 Tzs att åka in till auktionsmarknaden i Zanzibar Town för sälja sina frukter. Eftersom efterfrågan från hotell och restauranger är mycket hög säljer hon dagligen slut på all sin frukt.

Beräkningarna ska redovisas och svaren tydligt framgå.

- A. Definiera dina slumpvariabler och använd dessa för att skriva ett uttryck för (formel) Mwanajumas vinst.
- B. Beräkna medelnettovinsten.
- C. Beräkna standardavvikelsen för nettovinsten.
- D. Vilka (om några alls) antaganden behöver du göra för beräkningarna i deluppgift B och C?

Uppgift 4

(4 p)

Livslängden för glödlampor av ett visst märke är en normalfördelad variabel med väntevärde 1 200 timmar och standardavvikelsen 50 timmar. Ur produktionen väljer man slumpmässigt 25 lampor.

Beräkningarna ska redovisas och svaren tydligt framgå.

Hur stor är sannolikheten att medellivslängden understiger 1 180 timmar?

Uppgift 5

(4 p)

"CPMP-metoden ökar provpoäng med i genomsnitt 6,5 till 12,5 poäng jämfört med traditionell undervisningsmetod generellt sett med 95 % säkerhet."

Df= 577

Beräkningarna ska redovisas och svaren tydligt framgå.

Beräkna

- A. Felmarginal (margin of error).
- B. Punkttestimat, dvs medelvärdesskillnaden i stickprovet

Uppgift 6

(6 p)

Redogör för five-number-summary och boxplot genom att ge ett exempel och diagram (rita noggrant). Redogör strukturerat och tänk på att exemplet och diagrammet ska vara trovärdigt.

Uppgift 7

(6 p)

Studera utskriften nedan och besvara följande frågor. Pris på hus i \$1000-tal och boyta i kvadratfot.

```
Dependent variable is: Price
1000 total cases
R squared = 62.43%
s = 57930 with 1000 - 2 = 998 df
Variable      Coefficient
Intercept     6378.08
Living Area   115.13
```

- A. Skatta huspris för hus med boytan 260 kvadratfot.
- B. Vad säger lutningskoefficienten i just denna kontext?
- C. Vad säger R-square värdet i just denna kontext?
- D. Beräkna r.

Uppgift 8

(6 p)

Studera den första korstabellen (ANTAL) och besvara frågor som följer.

Det räcker om du redovisas uträkning av ETT procenttal per tabell vid sidan av respektive tabell.

1. ANTAL	Kvinna	Man
Ja	165	115
Nej	80	140

A. Fyll i tabellen med **radprocent** och skriv en tolkning av ett (och endast ett) procenttal.

2. RADPROCENT	Kvinna	Man
Ja		
Nej		

Tolkning A: (Skriv på linjen)

B. Fyll i tabellen med **kolumnprocent** och skriv en tolkning av ett (och endast ett) procenttal.

3. KOLUMNPROCENT	Kvinna	Man
Ja		
Nej		

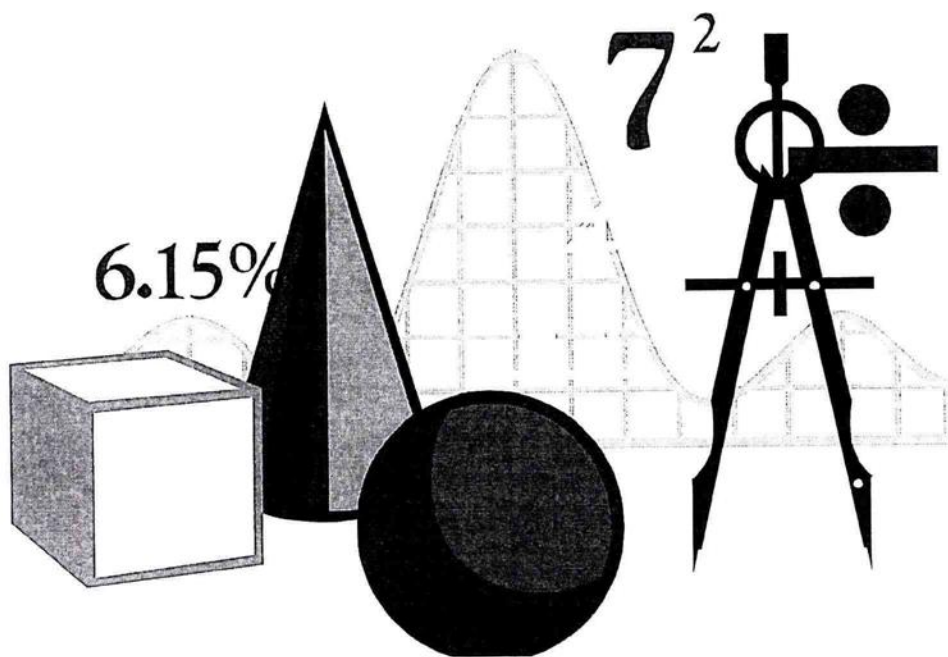
Tolkning B: (Skriv på linjen)

C. Fyll i tabellen med **totalprocent** och skriv en tolkning av ett (och endast ett) procenttal.

4. TOTALPROCENT	Kvinna	Man
Ja		
Nej		

Tolkning C: (Skriv på linjen)

REDOVISA UTRÄKNINGAR OCH SVAR PÅ DETTA BLAD SOM DU RIVER LOSS OCH LÄMNAR IN.



Formelsamling i statistik, A1, 15 hp

2016-01-25

Innehåll

BETECKNINGAR	2
KAP 1 DATA AND DECISIONS	2
KAP 3 DISPLAYING AND DESCRIBING QUANTITATIVE DATA	2
KAP 4 CORRELATION AND LINEAR REGRESSION.....	3
KAP 5 RANDOMNESS AND PROBABILITY.....	3
KAP 6 RANDOM VARIABLES AND PROBABILITY MODELS.....	3
KAP 7 THE NORMAL DISTRIBUTION.....	4
KAP 9+11 SAMPLING DISTRIBUTIONS.....	4
KAP 9-15 INFERENCE.....	5

Beteckningar

Stora bokstäver, X, Y etc., betecknar slumpvariabler.

Små bokstäver, x, y etc., betecknar faktiska värden på observationer.

Populationsstorleken betecknas med N .

Stickprovsstorleken betecknas med n .

Grekiska bokstäver betecknar populationens parameter. Ex σ , som betecknar populationens standardavvikelse

Latinska bokstäver betecknar skattningen av parametern. Den skattas utifrån stickprovet. Ex s , som betecknar stickprovets standardavvikelse.

	<u>Parameter (i populationen/ sannolikhetsfördelningen)</u>	<u>Parameterskattning (statistika) (i stickprovet)</u>
Medelvärde	μ	\bar{y} eller \bar{x}
Varians	σ^2	s^2
Standardavvikelse	σ	s
Andel, proportion	p	\hat{p}
Generellt	θ	$\hat{\theta}$

Kap 1 Data and Decisions

Five W' + How: Who, What, When, Where, Why and How

Kap 3 Displaying and Describing Quantitative Data

$$\text{Range} = \text{Max} - \text{Min}$$

$$\text{IQR} = Q3 - Q1$$

$$\text{Outlier Rule-of-Thumb: } y < Q1 - 1.5 \times \text{IQR} \text{ or } y > Q3 + 1.5 \times \text{IQR}$$

$$\bar{y} = \frac{\sum y}{n}$$

$$s = \sqrt{\frac{\sum (y - \bar{y})^2}{n - 1}}$$

$$z = \frac{y - \mu}{\sigma} \text{ (model based)}$$

$$z = \frac{y - \bar{y}}{s} \text{ (data based)}$$

Kap 4 Correlation and Linear Regression

$$\hat{y} = b_0 + b_1 x.$$

$$e_i = y_i - \hat{y}_i$$

Kap 5 Randomness and Probability

- The **Complement Rule** says that $P(A) = 1 - P(A^c)$.
- The **Multiplication Rule** for independent events say that $P(A \text{ and } B) = P(A) \cdot P(B)$, provided A and B are independent.
- The **General Multiplication Rule** says that $P(A \text{ and } B) = P(A) \cdot P(B|A)$.
- The **Addition Rule** for disjoint events says that $P(A \text{ or } B) = P(A) + P(B)$, provided events A and B are disjoint.
- The **General Addition Rule** says that $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.
- Events A and B are **independent** if $P(B|A) = P(B)$.

Kap 6 Random Variables and Probability Models

Know how to find the mean, or expected value, of a discrete probability model and the standard deviation:

$$E(X) = \mu = \sum x \cdot p(x)$$

$$SD(X) = \sigma = \sqrt{\sum (x - \mu)^2 \cdot p(x)} \text{ alternativt}$$

$$\sigma^2 = \text{Var}(X) = \sum (x - \mu)^2 \cdot P(x)$$

$$\sigma = SD(X) = \sqrt{\text{Var}(X)}$$

Foresee the consequences of shifting and scaling random variables:

$$E(X \pm c) = E(X) \pm c$$

$$E(aX) = aE(X)$$

$$\text{Var}(X \pm c) = \text{Var}(X)$$

$$\text{Var}(aX) = a^2 \text{Var}(X)$$

$$SD(X \pm c) = SD(X)$$

$$SD(aX) = |a| SD(X)$$

Understand that when adding or subtracting random variables the expected values add or subtract well:

$$E(X \pm Y) = E(X) \pm E(Y).$$

However, when adding or subtracting independent random variables, the variances *add*:

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$$

$$\text{SD}(X \pm Y) = \sqrt{(\text{SD}(X))^2 + (\text{SD}(Y))^2} \text{ if } X \text{ and } Y \text{ are independent}$$

Binomial Model for Bernoulli Trials: Binom (n, p)

n = number of trials

p = probability of success (and $q = 1 - p$ = probability of failure)

X = number of successes in n trials

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \text{ where } \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Mean: $\mu = np$

Standard deviation: $\sigma = \sqrt{npq}$

Kap 7 The Normal Distribution

Standard Normal distribution, $N(\mu=0, \sigma=1)$

$$z = \frac{y - \mu}{\sigma}$$

The Normal Approximation for the Binomial

A discrete Binomial model with n trials and probability of success p is approximately Normal if we expect at least 10 successes and 10 failures:

$$np \geq 10 \text{ and } nq \geq 10$$

The Normal distribution to use will have the following mean and standard deviation:

$$\mu = np$$

$$\sigma = \sqrt{npq}$$

Kap 9+11 Sampling Distributions

Model the variation in statistics from sample to sample with a sampling distribution.

- The sampling distribution of the sample proportion is Normal as long as the sample size is large enough.

Understand that, usually, the mean of a sampling distribution is the value of the parameter estimated.

For the sampling distribution of \hat{p} , the mean is p .

The standard deviation of the sampling distribution of a

proportion is $SD(\hat{p}) = \sqrt{\frac{pq}{n}}$ where $q = 1 - p$.

Sampling distribution of \bar{y} :

(CLT) As n grows, the sampling distribution approaches the Normal model with

$$\mu(\bar{y}) = \mu_y \quad SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$$

Kap 9-15 Inference

Inference:

Confidence interval for parameter = *statistic* \pm *critical value* \times *SE(statistic)*

$$\text{Test statistic} = \frac{\text{statistic} - \text{parameter}}{SD(\text{statistic})}$$

Eller

$$\text{Test statistic} = \frac{\text{statistic} - \text{parameter}}{SE(\text{statistic})}$$

Parameter	Statistic	SD(statistic)	SE(statistic)
p	\hat{p}	$\sqrt{\frac{pq}{n}}$	$\sqrt{\frac{\hat{p}\hat{q}}{n}}$
μ	\bar{y}	$\frac{\sigma}{\sqrt{n}}$	$\frac{s}{\sqrt{n}}$
$\mu_1 - \mu_2$	$\bar{y}_1 - \bar{y}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
μ_d	\bar{d}	$\frac{\sigma_d}{\sqrt{n}}$	$\frac{s_d}{\sqrt{n}}$
β_1	b_1	(in simple regression)	$\frac{s_e}{s_x \sqrt{n-1}}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	-----	$\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$

